

PT Novel methods and kits for detecting the presence of *Bacillus anthracis*
 PT In a test sample -
 XX
 PS Claim 2; Page 59-60; 62pp; English.
 XX
 CC The present sequence represents a bacterial surface array protein (SAP).
 CC SAP is used in the method of the invention. The specification describes
 CC a method for detecting the presence of *Bacillus anthracis* in a test
 CC sample. The method comprises contacting the sample with a capture
 CC reagent and detecting whether the a SAP is bound to the capture
 CC sample, which is indicative of the presence of *B. anthracis* in the
 CC sample. The method is useful for detecting the presence or absence of
 CC *B. anthracis* in a test sample.
 XX
 SQ Sequence 785 AA:

Query Match 100.0%; Score 3885; DB 22; Length 785;
 Best Local Similarity 100.0%; Pred. No. 2.1e-203;
 Matches 785; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGKTPDVPADHMGIDISINYLVKGA VGNKGNDKGFEPGKELTRAAATMAOITLMPIDK 60
 DB 1 agktfpdvpadhmgidsinylvkga vgnkgndkgmfepegkelttraaatmaoiltlmpidk 60
 QY 61 DAKPSFADSGQWYTPFTAAVEKAGVIRKGTGNGFEPNGKIDRVSMASLVEAYKLDTRKVN 120
 DB 61 dakpsfadsqgwytpftaaavekagvirkgtingfepngkidrvsmasllveaykldtrkvn 120
 QY 121 GTPATKFKDLETUNMGKREANITVELGTSVGTGDDQWEPKRYTKKAAOFTAKTDKQGT 180
 DB 121 gtpatkfkldletunmgkrehanitvelgtsvgtdqwpkrytkkkaaoflaktdkqgt 180
 QY 181 EAAKVESAKAAYTQKVEVVKFSKAVEKELTKREDIKVTNKANNDKVLKVEYTLSEDKRSATVE 240
 DB 181 eakvesakayvtqkvevfkfavekklkedikvtnkannndkvlkveytlsedkrsatve 240
 QY 241 LYSNLAAKOTYTVDNKKYKGTENVANGSLKAEATIEKADOTVVADEPTALQPTVKDENGTEV 300
 DB 241 lysnlaakqlycvdnkkygktnvngslkvaagsleaktiemadqlvvaadepalqftvkdengeev 300
 QY 301 VSPGIEFVTPAAERINAKGETITLAKGTSTTVKAAVYKKDKGRVVAESKREKVSAGCAAVAS 360
 DB 301 vspgiefvtpraaeklnakgetitlakgststtvkavyykkdkgrvvvaeskrekvsaagcaavas 360
 QY 361 ISNMTVAEQNKADFTSKDFKONKNVYEGDNAYVOVELKDQFNNAVTTGKVEYESLNTVEAY 420
 DB 361 isnmtvaeqnkadftskdfkgnknkyegdnayvvelkdqfnavttgkveyeslntevay 420
 QY 421 VDKATGKVTYLSAGKAPKVTYVKDSKALVSHVETEARFQKMKDITLKTVALSTK 480
 DB 421 vdkatgkvtvlsagkapkvtvkskalkvshvetearfqaqmkdiklktvalstlk 480
 QY 481 DYTDLKAKAPVLDQYKEFTAPVTYKVLDRDGELEKQKEAKYVNNRELVNNAAGQAGN 540
 DB 481 dvtldkavpvlldyqkeftapvtvkvldrdgkelkqkkaeyvnnrelyvnaagqagn 540
 QY 541 YTYVLTAKSSEKAKATLAEKAPGAFSEFVRGDLTBLDKVTVEBNQNMAMTVSLVP 600
 DB 541 ytyvltakssekaekatlalpakgafsfefvrgldteldkvyteenqnmamtvslvp 600
 QY 601 DANGLVYKGAEEAELEKTTTNKSGKEVDATDAOVTONNSVITVGGAKRGERTYKVTYVL 660
 DB 601 danglvlykgeaeaelktttnksgkevdatdaovtvnnsvltvggagakecykvtvyl 660
 QY 661 DGLLITTHSKRVVDTPATKAGLAVEFTSTLKEVAPNADLKAALLILSDVGPATTTAKA 720
 DB 661 dgllitthskrvvdtpatkaglaveftstlkevapnadlkaallilsvdgpatttaka 720
 QY 721 TAANVFVSADTVVAENGVGAKGATSTYVKULTVYKKGEGKGVFEFDKAVOAVASIRKA 780
 DB 721 taanvfvsadtlvvaengvlgakgatsiyvknltvvkdgkegkvfedkavavaasirkea 780

QY 781 KPATK 785
 DB 781 kpak 785

RESULT 2

AAW22863
 ID AAW22863 standard; Protein: 921 AA.

AAW22863;

27-FEB-1998 (first entry)

Bacillus steaerothermophilus s-layer protein sbs-B.

s-layer; sbs-B; vaccine; adjuvant; carrier; hybridisation assay;
 molecular spinning nozzle; molecular laser.

Bacillus steaerothermophilus.

Key	Location/Qualifiers
Peptide	1..31
Peptide	/label= sig-peptide
Peptide	32..921
Peptide	/label= mat-peptide

DE19603649-A1.

07-AUG-1997.

01-FEB-1996; 96DE-1003649.

01-FEB-1996; 96DE-1003649.

(LUBI/) LUBITZ W.
 (SLEY/) SLEYTR U.

Kuen B, Lubitz W, Sleytr U;

WPI: 1997-394558/37.

DR N-PSDB; AAT75488.

Preparation of s-layer proteins by expressing sbs-A gene in Gram
 negative bacterium - or new sbs-B gene in any host, also new
 recombinant proteins containing heterologous inserts, e.g.
 epitope(s), useful as vaccines and adjuvants

Claim 26; Pages 19-23; 31pp; German.

The present sequence is the *Bacillus steaerothermophilus* PV72
 s-layer protein, sbs-B. s-layer structures can be used as vaccines
 or adjuvants, particularly when they include a bacterial ghost that
 may contain additional epitopes in its membrane. Other uses of
 recombinant sbs-B, depending on the nature of the inserted peptide,
 are as an universal carrier for biotinylated reactants for use in
 immunological or hybridisation assays (the insert is streptavidin),
 to induce immune responses (epitopes), as a reagent for removing
 cytokine or toxin from serum (antigenic epitopes), as a molecular
 spinning nozzle (polyhydroxybutyrate synthase) and as a molecular
 laser (luciferase).

Sequence 921 AA:

Query Match 10.8%; Score 419.5; DB 18; Length 921;
 Best Local Similarity 24.0%; Pred. No. 4.8e-15;
 Matches 211; Conservative 133; Mismatches 332; Indels 203; Gaps 40;

QY 4 TFPDVPADHMGIDISINYLVKGA VGNKGNDKGFEPGKELTRAAATMAOITLMPIDKAK 63
 DB 33 sftdpvapy--kdaadflivsgatkgctekfygyddelttridaavilaviklvdnakd 90
 QY 64 PSFADSGQWYTPFTAAVEKAGVIRKGTGNG--FEPNGKIDRVSMASLVEAYKLD----- 116

Db 480 sfeytfxdn-----tkrtftvdvkkdpylgdtkvdatstvklsdeavvggeevgnqklix 534
QY 489 APVLDQYKEF-----TAPVTYVLDKQ-----GKELKEQ----- 518
Db 535 vsaodqgkkelkftgkykvrtnteglviknvnndntldfsgnsatcdqfvvaatkxi 594
QY 519 --KLEAKYVRELVLNAGOGAGNYTVLTAKEGSEKAKAT-LALELKAPGAFSEFEVR 574
Db 595 vngkevvyfxn-----asdtptstktlcnvvnvkadatpvgldivap-----seidvn 645
QY 575 GLDDELKYYTEENOKNAMYVLPVDANGVLKGAELKVTYTTNKEGKEVDATDQV 634
Db 646 apnt---astadvdfinfevevlyclidsngnrIk-----kvpt---attlvgtndyv 692
QY 635 TVONNSVTVTCOG-----AKAGETVYKVTVVLDOKLITTH-SFKVVDTPATKAGLAVEFTST 689
Db 693 evnngvniqfkgndeltltsstsvnvdtadg--ltkrlipkylnsaavpasatvratp 750
QY 690 SLKEVAPNADIKALNLITLSYDGPATTAKATASVNEFEVSADFNVAENGTVGAKGATSI 749
Db 751 tvklinsndltfeeltfgvid--ptqlvkdedln-eflav-----skaakndgyl 798
QY 750 YVKNLTIVYKD 759
Db 799 ynkplvtvkd 808
RESULT 5
AAR80530 ID AAR80530 standard; Protein: 1252 AA.
XX AAR80530;
AC
XX
DT 22-DEC-1995 (first entry)
XX
DE B. sphaericus SLP.
XX
KM Surface layer protein; SLP; fusion protein; vaccine; antigen;
XX surface expression; epitope.
XX
OS Bacillus sphaericus.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
XX
PN MO9519371-A2.
XX
PD 20-JUL-1995.
XX
PE 13-JAN-1995; 95WO-EP00147.
XX
PR 14-JAN-1994; 94GB-0000650.
XX
PA (SOLV) SOLVAY SA.
XX
PI Deblaere RY, Desomer J, Dhese P;
XX
DR WPI: 1995-263827/34.
DR N-PSDB: AAQ99430.
XX
PT Host cell expressing surface layer protein fusion protein - used for
PT host presentation of antigens and vaccine prodn.
XX
PS Disclosure; Fig.6; 95pp; English.
XX
CC A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
CC P-13855) surface layer protein was used to screen an HindIII-
CC generated library to isolate the slp gene. Promoter regions
CC of the gene are used in genetic constructs providing surface
CC expression of heterologous proteins in P-1 hosts.
XX

SQ Sequence 1252 AA:
Query Match 9.6%; Score 372.5; DB 16; Length 1252;
Best Local Similarity 22.5%; Pred. No. 25-12;
Matches 227; Conservative 129; Mismatches 373; Indels 281; Gaps 46;
QY 16 DSINLYEKGAVGKNDKMFEPGKELTPRAEAATMAQILNLPIDKAKPSPAD-SOGQWY 74
Db 45 eavgalvdgylgdtngnfnplntvtragaeflkealeangdv--dfkdkvkgawy 102
QY 75 TPPTAAVKAQVING-TGNGEFPNGKIDRVSMASLIVPAKLVLDIKVNTPTATKROLETL 133
Db 103 ynslaaavanglfigvsatefapnkslrsaaakllveafiglegad--lsetadassqv 159
QY 134 N-WGKEKANILVEIGISVGT-GDOWEPKVTYTKEAQFIKAT-DKQGTAAKAVESAKA 190
Db 160 kpwaakkyieavanglfigetdanklnpnmslitrqdfalvfrtvdkevegdepeeaafvka 219
QY 191 VTQKVEYKFAVEKLTAK-----EDIKVTN--KANNDKVLV-----KEVTLSED 233
Db 220 inntvevtfeeevtvngalnflklegleiknasvqtnkvvltteaqfdekeyvltld 279
QY 234 KRSATVELYSNLAQKQYTVVNVKNVGTREVAV-GSLEAKTITEMADQYVAADEPTA---LQ 289
Db 280 --geligfigkvaa---vptkvelvssavqklggevkvqakvtvaeqsgskaglpvt 332
QY 290 FTVKDENGTEVVS-----PEGIEFTVTPAA----- 313
Db 333 ftvpgnmndgyvprltlgealnneegiatystrykegidevtayatactdrsfsglyvfwg 392
QY 314 -EKINAKGEIT---LAKGTSTYKAVYK--KDGKV-----VAESKEVYV 351
Db 393 vdlisveevltggaavnganktykvknpkctgpeakntffngfyeanmnvtadkvana 452
QY 352 SAEGAAVASISNMWYVAEONKADFTSKDEKONNKVYEGONAVQVLELKQ-----Q 400
Db 453 tvngvkalqinsngald--aaqiltidskgeatftvsgtnaaavtpvpyllstnnstnkk 510
QY 401 FNA-----VTGKVEYESINTEVAV---VDKATGKVTVLSAGKA-PVVFYVYKD----- 444
Db 511 ysaalqgtlaskvftfaalqaeyletletradna-gevaalgtntngreykvlvckhavanlak 569
QY 445 -----SKGKALVSHITVEIEAF-----AKAKADIKLEETN----- 474
Db 570 neivnvaftedkdrvisvttnakfyvtdqdcavayftgdkk-kqlsv-ktbdkgeatftig 627
QY 475 -----VALSTKDVNTDLKAKAPVLDQYKEFTAPVTV-VVLDKDGKELKEOKLEAKYV 525
Db 628 sdtvndyapciawidintsdkqglddegepkavapisyfgapyldgsaalkayk---ksd 684
QY 526 NRELVLNAGOGAGNYTVLTAKEGSEKAKATLALAKPAGAFSEFEVRISDTELDKYYT 585
Db 685 lmkavtkfdgetavfaaelvngsgkvrtgtsl-----kkatyctlyngbandikvndqvl 739
QY 586 EENOKNMTVSVLPVDANGVLKGAELAE-----LKVYTTT-----KKEGKEVDATD 631
Db 740 spnsysvtlyea-tlssctgtvltlpaknlevtsvdgkttavvlatgtlavpddgkydyafta 798
QY 632 AQGVY-----QNNSVIIVGO-----GAKAGETVYK----- 655
Db 799 keatatftatnevpnsygtvatqfntadsqnsnslwafagknpykyaavgsqkykyfgan 858
QY 656 -----VTVLDGKLTTHSPFVVDTPATKAGLAVEFTST 690
Db 859 gnevfgesaealltqyategkvltisynvgdvtc---fkvis-----avnsstea 907
QY 691 LKEVAPN--ADLKAALNLITLSYDGVPAFTTAKATASVNEFEVSADTNVYA---ENGTVGAK 744
Db 908 ikpvaoptpaaprtfgaltltpaagglvltatcattlgtlsladadnvsatltvdcavtslk 967
QY 745 GATSIYVKNLTIVLVGDKGKQKVEFPDKAVQ-----VAVSIKEAKPA 783

Db 968 dsannsl-bltivetgantgv-factvgaagcllsltagcltltvtyadakna 1015

RESULT 6
AAR77673
ID AAR77673 standard; Protein; 1228 AA.
AC AAR77673:
XX
DT 16-JUL-1996 (first entry)
XX
DE S-layer protein encoded by sbsa gene.
XX
KM sbsa gene: S-layer; cell surface layer; expression.
XX
OS Bacillus stearothermophilus.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
FT Protein 31..1228
FT /label= mat_peptide
XX
PN DE4425527-A1.
XX
PD 25-JAN-1996.
XX
PF 19-JUL-1994; 94DE-4425527.
XX
PR 19-JUL-1994; 94DE-4425527.
XX
PA (VOGE-) VOCLEBUSCH GMBH.
XX
PI Lubltz W;
XX
DR WPI; 1996-077933/09.
DR N-PSDB; AAT08695.
XX
PT Nucleic acid encoding signal peptide of Bacillus stearothermophilus
PT S-layer protein - which has a lysine content of at least 10 per
PT cent.
XX
XX Disclosure: Page 6-7; 12pp; German.
XX
PS The claimed signal peptide encoding sequence (see tag e of AAT08695)
CC is pref. operably linked at the 3' end to a protein encoding
CC sequence. The protein is pref. the S-layer protein (see tag f of
CC AAT08695). At the 5' end, the signal peptide encoding sequence is
CC pref. linked to an expression control sequence, pref. the sequence
CC given in AAT08696.
XX
SQ Sequence 1228 AA;

Query Match 6.4%; Score 248.5; DB 17; Length 1228;
Best Local Similarity 22.4%; Pred. No. 1.3e-05;
Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;

OY 55 NLPTRKDAKPSFADSGOGMTYTFPIAIVEAGYIKGTGNGFEPNGRIDRVSNASLLVEAYK 114
DB 323 dvlsltdgkltlvdas---tpfennteykvvvgkik--dkngkefkedaftfkrlnda 375
OY 115 LDTKVNCTPAT-----KFKDLETLN-----WGK----- 137
DB 376 vltgvtgtnvntvsnlaagltfdtdcltlvfdkllepelvsnvtiltdvetygkripv 435
OY 138 ----EKANILVELGISVGTGDQME---PKRTVT--KAEAOFIAKTDKOFTTEAKVES 187
DB 436 latsgstlttlkcalvltgkykialnvnkltlgyneayelv-----ftanasaptv 489
OY 188 AKAVTT-----OKVEVFSKAVEKLTREDIKVT 215
DB 490 ataptlgttltstlgtltnwvgklagvneaglytgp9lqfttltfatkldestladnftvl 549

OY 216 NKANNDKVLKVEVTLSEDKRSATVELYSLNAKQTYVDNVKVGKTE--VAVGSLEAKTI 273
DB 550 vekesgtvvaselkynadakmvtlvpkadlkentllykikikglsdkgielgtvnekly 609
OY 274 EMADQTVVADDEPTALOFTVKD-ENGTEVVSPEGIEFVTPAAEKINAKGSETTLAKGST-- 330
DB 610 eefktgdlta--plvtlsvtskngdaglkvt--eaqetfvfsenlnftnatlvsgstllyg 665
OY 331 -----TWKAVYKKDG---KVVAESKE----- 348
DB 666 qvavvkgaganlaltasdlipasevaevtgqdgtykvkvaanglerngykllvfgkgata 725
OY 349 -VKVSAEAGAAVASISNMTVAEQNKADFTSKDKRKNKKVVEGONAVQVELKQFPAVNTTG 407
DB 726 pykdaaanatlalnyiyltltteqg-dvtaplv--tkvylgds-----lkda-davlt- 773
OY 408 KVEYESLNTAEVAVVDKATGKVTVLSGKAPVKVTV-----KDSGKALVSHTEIEAFA 461
DB 774 -----lnvdaqg---kftlqfseelktssg-slyvgkvtevekl 809
OY 462 OKAMDKIEKTNVALSTVDVTLKVKAPVLDQYGRKEFTAPVTVKVLDDGKELEQKLE 521
DB 810 nngwvd---agtgctevsv-----apktadangkvtlaavvtlqldmndkdak-rliv 856
OY 522 AKVNRRELVLNNAQO-----EAGNTYVVLTAKSGEKEAKATLALDELKAPGAFSK 570
DB 857 vksstgdagdvaagnvikedlllynswrthvasvkaadqg-----naaaatlp- 908
OY 571 FEVRGLDTELD--KYTEENOKNMTVSVLPVDANGVLKGAEAELKVTYTNKEGKEV 627
DB 909 -----tstaiddtksljvefine-----ldlaevkpenlivvdaagnavagvtaldgs- 956
OY 628 DATDAQVYQNNSVYTVGOGAKAGETKYKVTVLDG-----KLITTSFKVVDTPAP 677
DB 957 -----tnkfvtlpsqelkagltvsvtl--dgvrckvgnltskylt--sfktvsanp 1003
OY 678 TAKGLAVEFTSTLKEVAPNADLKAALNLSVDDVP-ATTAKATASWVERFSAD-TWVY 735
DB 1004 t-----lsslsiadgavvnd-rsktltliefsdsvpnpptlllkadgstltllyclvnm 1055
OY 736 AENGT---VGAKGAT--SIYKKNLTVVKDGK----EOKVEPDKAVOVAYSIRKEAKPA 783
DB 1056 nenklykivfhkyvltdefltyelavaskdftgtdidaskvtf---ltsvatdevkpa 1110

RESULT 7
AAM22862
ID AAM22862 standard; Protein; 1228 AA.
XX
AC AAM22862;
XX
DT 27-FEB-1998 (first entry)
XX
DE Bacillus stearothermophilus S-layer protein sbs-A.
XX
KM S-layer; sbs-A; vaccine; adjuvant; carrier; hybridisation assay;
KM molecular spinning nozzle; molecular laser.
XX
OS Bacillus stearothermophilus.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= sig_peptide
FT Peptide 31..1228
FT /label= mat_peptide
XX
PN DEL9603649-A1.
XX
XX 07-AUG-1997.
XX
PF 01-FEB-1996; 96DE-1003649.

PR 01-FEB-1996; 96DE-1003649.
 XX (LUBI/) LUBITZ W.
 PA (SLEY/) SLEYTR U.
 XX
 PI Kuen B, Lubitz W, Sleytr U;
 XX
 DR WPI; 1997-394558/37.
 DR N-PSDB; AAT75487.
 XX
 PT Preparation of S-layer proteins by expressing sbs-A gene in Gram
 PT negative bacterium or new sbs-B gene in any host, also new
 PT recombinant proteins containing heterologous inserts, e.g.
 PT epitope(s), useful as vaccines and adjuvants
 XX
 PS
 XX Claim 1; Pages 9-14; 31pp; German.
 CC The present sequence is the Bacillus stearothermophilus PV72
 CC S-layer protein, sbs-A. S-layer structures can be used as vaccines
 CC or adjuvants, particularly when they include a bacterial ghost that
 CC may contain additional epitopes in its membrane. Other uses of
 CC recombinant sbs-A, depending on the nature of the inserted peptide,
 CC are as an universal carrier for biotinylated reactants for use in
 CC immunological or hybridisation assays (the insert is streptavidin),
 CC to induce immune responses (epitopes), as a reagent for removing
 CC cytokine or toxin from serum (antigenic epitopes), as a molecular
 CC spinning nozzle (polyhydroxybutyrate synthase) and as a molecular
 CC laser (luciferase).
 CC
 XX
 SO Sequence 1228 AA.

Query Match 6.4%; Score 248.5; DB 18; Length 1228;
 Best Local Similarity 22.4%; Pred. No. 1.3e-05;
 Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;

QY 55 NLRIDDAKSFADSOQWTPPIAAVEKAGVIRKGTNGCEPRKIDRVMSASLVATK 114
 DB 323 dvalstgdkltvda---tpienteykvvvkylk---dkngkefkeadaftklinda 375
 QY 115 LDRKVNQTPAT-----KFKLETLN-----WGK-----137
 DB 376 vvtgvtgntvntmtsvnaagftddltlvridklapetvnsnvtldvetgkripv 435
 QY 138 -----EKANILVELGISVGTGDQWE---PKRYVT--KAERAOPIAKTDKQFGTEAAKVES 187
 DB 436 iastsgftitlkealvtgkylainnvkltlgynaeayelv-----ftanaaptv 489
 QY 188 AKAVTT-----QKVEKFSKAVKRLKREDIKVT 215
 DB 490 ataptlgtltsltslctnvwglagvnaegtlyypjlgfttfackldestladnflv 549
 QY 216 NKANNDKVLKEVTLSEDKSARFVELSNLAQOTYVDNVKVGKTE--AAVGSLEAKTI 273
 DB 550 vkesgtvsvselkynadadkmvltvpkadikentlgyiklkglsksgleigtvneky 609
 QY 274 EMADQIVVADEPALQFTVKD-ENGTEFVSPGIEFVTPAAEKINAKGEITLAKGTST-- 330
 DB 610 eftqgdltl-pvltsvtskngdaglkvt--eageftvcktsenlntfnatvsgstltyg 665
 QY 331 -----TYKANVKKDG-----KVVAESKE-----348
 DB 666 qvavkvaganlsaltaadltpasveavtgdgtlykvvnaanglerngykvllvfgkgata 725
 QY 349 -VWVSAGAAVVASISMTVAEONKADFTSKDFKONKNVGVGDNANVVOELKDOGNATYG 407
 DB 726 pvtadaaanlatlcnlylfttegg-dvtaptv---tkvifgds-----lkda-davtl- 773
 QY 408 KVEESILNTEVAVVAKATGKVTLSAGKAPVKTV-----KDSKGRALVSHTEIEAFA 461
 DB 774 -----ltnvdagq---kftfigfseelktsesg-slvsgkvvtvekl 809
 QY 462 QKAMKDIKLEKTVALSTKDVTDLKVAPVLDQYGKEFTAPVTVKVLDDKGKELKECKLE 521

DB 810 ngwvvd---agtlvtsv-----apktadangfvtaavvltglndndkxak-lliv 856
 QY 522 AKVYNRELVNNAAGQ-----EAGNYTVVLFAKSGEKEAKATTLALFLKAPGAFSK 570
 DB 857 vksscdgldadvagnvlkekdllilrnsrwhvsvkaadkdqg-----nasafp- 908
 QY 571 FEVRGLDTELD---KYVTEENOKNAAVTSLVPDANGVLKGAEAALKYTTTNNKEKEV 627
 DB 909 -----tstaiddtkslivefne-----tdlaevkpenlvkvdagaavagfvvaldgs-- 956
 QY 628 DARDQVTVONNSVITVVGAKAGERTYKVTYVLDG-----FLITTHSKKYVDFAF 677
 DB 957 -----tnkvtfipsqelkaglvsvrl--dgvrdkvgnltiskylt--stktvsanp 1003
 QY 678 TAKLAVEFTSTLKEVAPADILKALNLISVDGVP-ATTAKATASNVEFVSAD-TNVV 735
 DB 1004 t-----lssisladgavvvd-rsktltlefdsrpnptlcltkhdgsftnylvtnn 1055
 QY 736 AENGT---VGAKGAT--SIYKMLTVVKDOK-----EQKVEFDKAVOVAVSIEAKPA 783
 DB 1056 nenktykivfhkgvltideftqyelavskdftqtdidskvrlf---ltgsvaldevkpa 1110
 RESULT 8
 ID AAW93252 standard; Protein; 1228 AA.
 XX
 AC AAW93252;
 XX
 DT 24-AUG-1999 (first entry)
 XX
 DE B stearothermophilus sbsa protein.
 XX
 KW sbsa protein; S-layer protein; Gram-negative; prokaryotic host cell;
 KW integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
 KW eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
 KW immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
 KW molecular laser; universal carrier molecule; monomolecular layer.
 XX
 OS Bacillus stearothermophilus.
 XX
 XX
 FT Key Location/Qualifiers
 FT Peptide 1..30
 FT Protein /label= signal_peptide
 FT /note= "sbsa"
 XX
 PN DEL9732829-A1.
 XX
 PD 04-FEB-1999.
 XX
 PF 30-JUL-1997; 97DE-1032829.
 XX
 PR 30-JUL-1997; 97DE-1032829.
 XX
 PA (LUBI/) LUBITZ W.
 XX
 PI Lubitz W, Resch S;
 XX
 DR WPI; 1999-122189/11.
 DR N-PSDB; AAX22748.
 XX
 PT Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
 PT integrated into membranes or organelles or secreted into periplasma
 PT or growth medium, and nucleic acid encoding S-layer proteins with
 PT peptide insertions, used in vaccines or for enzymatic reactions
 XX
 PS Disclosure; Page 14-18; 34pp; German.
 CC This invention describes a method for the production of a S-layer protein
 CC (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
 CC transformed with nucleic acid (II) encoding (I), linked to a signal

CC sequence (SS) that encodes a protein which causes at least one of (1)
 CC integration of (1) into the external or cytoplasmic membranes and/or
 CC (1) secretion of (1) into the periplasmic space or extracellular medium,
 CC (b) culturing the cell to express (1) and (c) optionally recovering (1)
 CC from the membranes, periplasmic space and/or extracellular medium.
 CC Alternatively, a eukaryotic cell is used as host and then the SS, which
 CC is optional, promotes integration into the cytoplasmic membrane or an
 CC organelle and/or secretion into the extracellular medium. (1), and
 CC derived structures, may include a wide variety of polypeptide inserts or
 CC are useful as (1) vaccines or adjuvants (with immunogenic epitopes and
 CC immunostimulants inserts such as cytokines) (1) as reactors (inserts
 CC are enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a
 CC 'molecular spinnetete' for production of PHB or luciferase for use as a
 CC molecular laser (when combined with substrate and oxygen)) and (iii) as
 CC universal carrier molecule (streptavidin is inserted) for use in
 CC hybridisation and immuno assays, or for selective elimination of
 CC cytokines, toxins etc. from body fluids (inserts are specific binding
 CC epitopes). In this system, heterologous (1) do not form inclusion bodies
 CC but rather monomolecular layers, and in eukaryotic cells they undergo
 CC glycosylation. This sequence represents the Bacillus stearothermophilus
 CC sbSa protein which is used to illustrate the method of the invention.
 CC
 XX
 SQ Sequence 1228 AA;

Query Match 6.4%; Score 248.5; DB 20; Length 1228;
 Best Local Similarity 22.4%; Pred. No. 1.3e-05;
 Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;

OY 55 NLPIDKAKPSFADSGOGWYRPFIAAVEAGVIGKGTNGFBNQIDSVSMASLVEAKV 114
 DB 323 dvtstgdgkltlvdas-----tpfenteykvvvkgik---dangkefkdafitrlrnda 375
 OY 115 LDTKNGTTPAT-----KFKDLETLN-----MGK----- 137
 DB 376 vvtgyfglnvtnhtsvlnlaagltfdtdltlvfdklapetvsnsvltdvengkripv 435
 OY 138 ----EKANILVELGISVGTGQWE---PKRTVT--KKAQAQFAKTOKGTGEAKVES 187
 DB 436 lasstgsctltlkealvtgkylalnkvklltgyneaayelv-----ftanasapltv 489
 OY 188 AKAVTT-----QKVEVFSKAVEFTKEDIKVT 215
 DB 490 ataptcltgctlstgsltnwglagvneagtypglqcttctfatldesltladnfv 549
 OY 216 NKANDKYLVEVTLSEDKRSATVELYSNLAKOTYTVDNKVKGTE--VAVGSLEAKTI 273
 DB 550 vekesgltvvaselkynadakmvtlvpkadlkenltlygkikglskdjagltvnektly 609
 OY 274 EMADQTVVADPPTALOFTVKD-ENGTEVVSPEGIEFVTPRAEKINAGETILAGTST-- 330
 DB 610 ektcgdlta--ptvsvtskngdaglkvt--eageftvktfsenltnatvsgstltlyg 665
 OY 331 -----TVKAVYKKKG-----KVVAESKP----- 348
 DB 666 qvavvkaagauisatlasdlpasveavtggdgytkvkaanqlerngylvfgkdata 725
 OY 349 -VKVSAGCAVASTSNMTVAEQNAADFTSKDFKONKRVYECGDNAVVOVELKDQFNATV 407
 DB 726 pvkdaanaantltyltftlegg-dvtlaptv--tkvfygds-----lkda-davtlt- 773
 OY 408 KVEYESLNTFEAVVDKATGKVTVLSAGKAPVKYTV-----KDSKGALYSHTVEIFAFA 461
 DB 774 -----ltndagq---kftlfgseelktsfg--slvgkykvtekekl 809
 OY 462 QKAMKDIKLEKTNAVALSTKDVTDLKVKAPVLDQYGEFTAPVTVKVLDRKDGKELKEBGL 521
 DB 810 ngwvwd---agtgltvsv-----apktndngkvtlaavvltlgtddndkdak-rlly 856
 OY 522 AKVYNREVLVMAAGQ-----EAGNYTVVLTAKSSEKAKATLALLEKAPGAFSK 570
 DB 857 vdkssitdgladvagnvkkdkdlilrynswrhtvasvkaadkdqg-----nasafp- 908

OY 571 FEVRGLDTELD---KYVEENQKNMTVSVLPVDANGVLKGAEAELKVTTNKEGKV 627
 DB 909 -----tstaidctksllvefne-----tdlaevkpenlyvkkdaagnavagvtaltdgs-- 956
 OY 628 DATDAQVTVQNNSVITVVGOGAKAGETKYVTVVLDG-----KLITTHSFKVVNDAP 677
 DB 957 -----tnkfvtptspgelkagvtvsvlt--dgyrvdkvgnltkylt--sfktvssnp 1003
 OY 678 TTKGLAVEETSTSLKEVAPNADLKALNLSVDGV-ATTAKATASNVEFVSAD-TRNV 735
 DB 1004 t-----lssisladgavnd-rskltliefdsdvppltlkkadgstfngylvlnvn 1055
 OY 736 AENGT---VGAKGAT--SIYVKNLVVMDK-----EQKVEDKAVQVAVSIEKAKPA 783
 DB 1056 nenktykltvfhkgyvldefgtgylavskdfgtgltdidskvtf---ltgsvatdevkpa 1110

RESULT 9
 AABI0625
 ID AABI0625 standard; Protein: 1228 AA.
 AC AABI0625;
 XX
 DT 08-JAN-2001 (first entry)
 XX
 DE B. stearothermophilus sbSa protein.
 XX
 KW S-layer gene; sbSa; antibacterial; vaccine; adjuvant; bioreactor;
 KW polyhydroxyalkanoate synthesis.
 XX
 OS Bacillus stearothermophilus.
 FH
 FT Key Location/Qualifiers
 FT Peptide 1..30
 FT /label= signal_peptide
 FT Peptide 31..1228
 FT /label= mature_peptide
 PN
 XX DE19903345-A1.
 PD
 XX 03-AUG-2000.
 XX
 PF 28-JAN-1999; 99DE-1003345.
 XX
 PR 28-JAN-1999; 99DE-1003345.
 XX
 PA (LUBITZ) LUBITZ W.
 XX
 PI Lubitz W;
 XX
 DR WPI; 2000-533868/49.
 DR N-PDB; AAA71797.
 XX
 PT Host cell, useful e.g. as bioreactor for production of
 PT poly(hydroxyalkanoate), containing two or more recombinant
 PT polypeptides, with at least one in carrier-bound form -
 PS
 XX
 XX Disclosure; Page 12-16; 26pp; German.
 CC This invention describes a novel host cell (A) comprising at least two
 CC functional recombinant polypeptides (1), at least one being in carrier
 CC bound form. The products of the invention have antibacterial activity.
 CC (A), or, where bacterial, their ghosts (B), are useful as vaccines or
 CC adjuvants (specifically for presentation of immunogenic epitopes or
 CC pathogens or autologous immunostimulatory polypeptides, e.g. cytokines),
 CC or preferably, as enzyme reactors for performing a cascade of reactions,
 CC specifically synthesis of poly(hydroxyalkanoate). Localization of
 CC individual (1), specifically enzymes, in separate cellular compartments
 CC avoids adverse reactions between products and substrates, when being used
 CC as bioreactors. (1) can be produced in carrier-bound form without loss of
 CC function. This sequence represents the Bacillus stearothermophilus
 CC S-layer protein sbSa which is used to illustrate the method of the
 CC invention.

Matches 202; Conservative 122; Mismatches 342; Indels 281; Gaps 43;

DT	15-JAN-1997	(first entry)
XX		
DE	Haemophilus adhesion protein HA2.	
XX		
KM	Haemophilus adhesion protein; HA2; hsf protein; vaccine.	
XX		
OS	Haemophilus influenzae type b strain C54.	
XX		
PN	WO9630519-A1.	
PD		
XX	03-OCT-1996.	
PF		
XX	22-MAR-1996; 96WO-US04031.	
PR		
XX	24-MAR-1995; 95US-0409995.	
XX		
PA	(UNSL-) UNIV ST LOUIS.	
PA	(UNIM) UNIV WASHINGTON.	
XX		
P1	Barenkamp SJ, St Geme JW;	
XX		
XX	WPI; 1996-455364/45.	
DR		
DR	N-PSDB; AAT41476.	
XX		
PT	Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in	
XX	vaccines against H. influenzae infection.	
PS		
XX	Claim 5; Page 66-73; 120pp; English.	
CC		
CC	Haemophilus adhesion protein HA2 (AAR9393) is associated with the	
CC	formation of surface fibrils involved in adhesion to various host	
CC	cells; it is also referred to hsf (Haemophilus surface fibrils).	
CC	Its amino acid sequence was deduced from a genomic DNA clone	
CC	(AAT41476) derived from Haemophilus influenzae type b strain C65.	
CC	Large quantities of recombinant HA2 can be produced in transformed	
CC	prokaryotic or eukaryotic host cells, for use in vaccines against	
CC	H. influenzae infection.	
XX		
XX	Sequence 2353 AA;	

RESULT 11	
AAR9393	
ID AAR9393	standard; Protein: 2353 AA.
XX	
AC AAR9393;	
Db	357 tas-----vtkdtngngltvkvdkavgdglkf-----dsdskkl-----vadttaltv 399
	: : : : : : :
QY	334 AVYKKDGKVVAAESKE---VYVSAGCAAVASISN--WTVAEONKADFTSKDPKONNKVYEG 388
	: : : : : : :
Db	400 g-----gkvaeiakeddkkllvmagdlvtalgnswakakaadtgdgalegiskdqvkg 454

QY 389 DNAYVOVELKDQENAVTTGKVEYESLNTENVAVVDKATGKTV-----LSGKAPVKYTV 443
DB 455 etv-----tfkagknllvkqgdganfyslqdalgtlsitlgtltngndakvlnk 506
QY 444 D-----SKGKALVSHTVIEIAFAOKA-MKDIKLEKTNNVALSKVDYDLKVKAPVLDQY 495
DB 507 dglitltpagnggtlgtntlsivtkdglkagnka1-----tnvasglrayd----- 551
QY 496 GKEFTAPVTVKVLVDKQKELKEQ-----KLEAKYVNR-LVLNAGQEGAGNTYV-- 543
DB 552 -----anfvlmsatdlrnrvedaykglhlnneknankgprlvdstatvqddllk 604
QY 544 -VLTAKSGEKA--KATLALELKAPGAFSKFEVRGLDELDKYVTEENOKNAMTVSVLPV 600
DB 605 vsvtkngtkesngqvqadevltfga-----gaatvtsk---senghltlsvaet 654
QY 601 DANGVLKGAKEAEIKVTTNNKSGKEVDATDAQVTVQNNNSVITVGCAKGEFTYKVVVL 660
DB 655 kacgglekdgdllklyvdngn-----cdnylvlvgnn----- 685
QY 661 DGLITTHSKFVVDTAPPAKGLAVEFTSTSLKEVAPN-ADLKAAALLNLSVDGVPATTAK 719
DB 686 -glavtckgftetvktgat-----dadrgkvtkdatandackkvatvkdv-----atain 734
QY 720 ATASNNEFVSADTNVVAENGSTVGAK-----GATSIYV--KNTLVKDGKEQKVEFDKAV 771
DB 735 saatfvktenlttsidednptdngkdaalagadlttfkagknllkvkrdrgk--nltfdlak 792
QY 772 QVAVSIKEAK 781
DB 793 n--levktak 800

RESULT 12
AAB23860
ID AAB23860 standard; Protein; 2411 AA.
XX AAB23860;
AC AAB23860;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW non-typable Haemophilus influenzae; antiinflammatory; audiotory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen.
XX
OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000MO-CA00289.
XX
PR 16-MAR-1999; 9905-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loosmore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-61897/59.
DR N-PSDB; AAA92459.
XX
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -
XX
PS Claim 1; Fig 24; 275pp; English.
XX
CC The present sequence represents a Haemophilus influenzae adhesin
CC (Hia) protein from the type c Haemophilus influenzae strain API.

CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
SQ Sequence 2411 AA;

Query Match 6.1%; Score 238.5; DB 21; Length 2411;
Best Local Similarity 23.0%; Pred. No. 0.00011;
Matches 195; Conservative 102; Mismatches 333; Indels 219; Gaps 42;

QY 23 EKGAIVGNDKGMFEPGKELTRAEATMMQILNLPIDKXAK-PSFADSGQWTPPIA-A 80
DB 79 ekevtenmgwlyfhnkgyllkagaitlkagd-nlkxqstnaassfyslkldltlsva 137
QY 81 VERAGV-----IKGTGNGFE---PNGKI-----BRVMSASILVAVRYKIDFK 118
DB 138 teklsfngdkvdltsdangllaktngnyhnlngldstlpavcnhtglsstsfpn 197
QY 119 VNGTPATKFKDLETLNKGKKEANILVELGISVGTGDMQEPKVTYKAEOQTAKTDKQF 178
DB 198 vektraatvkdvlinaqvnllkgakt-----agvvesvdlsaynnveflltgdntll 248
QY 179 GTEAAKVESAKAYTQKVEKFSKAVEKLEKEDIKV-TNKAND--KVLKEXELTSLDKR 235
DB 249 dvvlatakengkt-----tevkftprktsvllkqdklltfgemndnlnvntatcdndeg 303
QY 236 SATVELYSNLAOKQTYTVDNKVG--KTEVA-----VGSLEAKTTEMAQDOTVAD 283
DB 304 nglvltakavida-----vnkagwrvlltlangngdftvasgnvltseagdtlas- 355
QY 284 EPTALOPTVVDENGTEYVSP-----EGIEFVTPPAEKINAKGETTLAKTSTTVKAVYK 337
DB 356 -----vltcdngnglvtkydakvgdglkf--dsdckl-----vadltaltvg-- 396
QY 338 KDKGVVAESKE---VKVASAGAAVASISNWTVAEONKADFTSKD-----FKONNKVYEGD 389
DB 397 --gkvaeiakeddckklvماغdlvltagnlswkakaead-tcdgagalegskdqvckage 453
QY 390 NAVVOVELKDQFNNAVTTGKVEYESLNTENVAVVDKATGKTV-----LSAKKAPVKYTVKD 444
DB 454 tv-----tfkagknllvkqgdganfyslqdalgtlsitlgtltngndakvlnkd 505
QY 445 -----SKGKALVSHTVIEIAFAOKA-MKDIKLEKTNNVALSKVDYDLKVKAPVLDQY 496
DB 506 glitltpagnggtlgtntlsivtkdglkagnka1-----tnvasglrayd----- 549
QY 497 KEFTAPVTVKVLVDKQKELKEQ-----KLEAKYVNR-LVLNAGQEGAGNTYV--- 543
DB 550 -----anfvlmsatdlrnrvedaykglhlnneknankgprlvdstatvqddllk 603
QY 544 VLTAKSGEKA--KATLALELKAPGAFSKFEVRGLDELDKYVTEENOKNAMTVSVLPV 601
DB 604 vsvtkngtkesngqvqadevltfga-----gaatvtsk---senghltlsvaet 653
QY 602 ANGLIVLKGAFAAEIKVTTNNKSGKEVDATDAQVTVQNNNSVITVGCAKGEFTYKVVLD 661
DB 654 adsglekdgdllklyvdngn-----tdnylvlvgnn----- 683
QY 662 GKLIITTHSKFVVDTAPPAKGLAVEFTSTSLKEVAPN-ADLKAAALLNLSVDGVPATTAK 720
DB 684 glavtvggftetvktgat-----dadrgkvtkdatandackkvatvkdv-----atains 733

```
QY 721 TASNVEFVSADTNVAENGTVGAK-----GATSIYV--KNLTVVKDGKFEQKVEFDKAVQ 772
D 734 aatfvktenlttsidednpndngkdalkegdltltfkaglnlvkrdgk--nltfdlaln 791
QY 773 VAVSIKFAK 781
D 792 --Levktak 798

RESULT 13
AAR12083
ID AAR12083 standard; protein; 1116 AA.
XX
AC AAR12083;
XX
DT 01-AUG-1991 (first entry)
XX
DE HWP protein.
XX
KM HWP; protein; food.
XX
OS Bacillus brevis HPD 31.
XX
FH Key
FT Peptide 1..53 Location/Qualifiers
FT Peptide 31..53 /label= sig_peptide
FT Protein 54..1116 /label= sig_peptide
FT Protein /label= mat_protein
XX
PN JP03094683-A.
XX
PD 19-APR-1991.
XX
PF 06-SEP-1989; 89JP-0229304.
XX
PR 06-SEP-1989; 89JP-0229304.
XX
PA (HIGE-) HIGETA SHOYU KK.
XX
DR WPI; 1991-159801/22.
XX
DR N-PSDB; AAQ11789.
XX
PT New protein HWP gene - comprising specified DNA sequences,
PS permitting produ. of large ams. of HWP for use as food protein
XX
PS Disclosure; Fig 1; 8pp; Japanese.
XX
CC Using the sequence encoding this protein, HWP can be produced in
CC large amounts by recombinant techniques. The protein is used in food.
CC The first amino acid is encoded by the triplet TTC, according to
CC the specification.
XX
SO Sequence 1116 AA.

Query Match 6.0%; Score 231.5; DB 12; Length 1116;
Best Local Similarity 21.6%; Pred. No. 0.0001;
Matches 193; Conservative 125; Mismatches 307; Indels 269; Gaps 47;
```

```
QY 190 AVTQKEVYKFSKAVEKITKEDIKVTNK---ANNDKVLVKEVTLSEDK--RSATVEIYKSN 244
D 213 -mlndalrvklnmeglegtdirlnvtdelltkylkvevrmndwehngnshelpvtn 271
QY 245 LAAQOTYVDVKNKVGTEAVAGSLEAKTTEMAADQTV--VAD--DEPTA-----LOFTYKDE 295
D 272 vpaiglskkanev-----clngkadlgsnltkvaeglnpnafdggkvgvwlkdd 323
QY 296 NGTEVYSEEGIEFVTPAAEKINA---KGEITLAKGTSTTVKAVYKKDKQVVAESKEVYKS 352
D 324 renvlyvmegsededvmdrvsalykkgaf-----tddivdlsksd-----lddvk 373
QY 353 AEGAVALASISNWTVAEONK-----ADFTSKDFKONN---KYEGDN--AYVQ 394
D 374 mdg-----seksyrltedtklcyntfrfdpvdalsklykndrtfgvkvIndnnevaylh 429
QY 395 VELKDQFNAVTTGKVEYESLNTTEVAVVDKATGKTVTVLSAGK----- 435
D 430 l-iddqtldksvkgvkygs--kvlskidadkkltnlndskfsdledqdegskdlvflgd 486
QY 436 APKVY-TVKDS-----KGAIVSHTVEIEAFQAQRAMDKLEKTNVALSTKDVY 483
D 487 qpaklgsdtkesdvsvyyadgdkkylv-----fanrvaeagkvek---vvsrntk 534
QY 484 DLKVKAPVLDQYGEKFTAPVTVKVLDDKGELEKQKIEAKTVNR--ELVINAAGQENAGY 541
D 535 dir-----lvvgk-----tkvyv-pdaasyseanankvkvksdldlslndgeev--- 579
QY 542 TVVTLTASGEKAKATLALELKAPGAF-----SKFEVRGIDTE 579
D 580 klldpsgrvrhietkdaiddrklplaitkgtatynskcdtydftmtgkgtivsltdqk 639
QY 580 -LDKYVTEENOKRAMVSVLPVDANGLVLKGAELAEIKYTTTNKSGKEVDATDAQVYQ 637
D 640 diyrtyvnyvdksndk-----rqaftexlveallqpkvksedsatdanqtv- 684
QY 638 NNSVTVVGQAGAKAGETTYKTVVLDGKL-----IT--THSFYV 672
D 685 ---llevnfsk-gevdkv-vldsklkysektwdkladeddvvgdyevtdaktavfkm 739
QY 673 V-DTAP---TAKGLAVEFTSTLSKEVAPNADL-----KAALNITSVDGVPATPTAK 719
D 740 tgdilpatgtlkgelknahtakfkdvakksdklvwysvdedkgevgatlvvgd-----s 793
QY 720 ATASNVEFVSADTNVAENGTVGAKGATSIYVKNLTVVKG-----KEQKVEFD 768
D 794 glvgdhqf-----gmvkqyglaskgdtlti-----vtkgdgsvelekeykldgd 836

RESULT 14
AAB15945
ID AAB15945 standard; protein; 2383 AA.
XX
AC AAB15945;
XX
DT 05-OCT-2000 (first entry)
XX
DE E. coli proliferation associated protein sequence SEQ ID NO:302.
XX
KM Escherichia coli; E. coli; proliferation; inhibition; screening;
XX antimitrobal; bacterial growth; antisense therapy; antibacterial.
XX
OS Escherichia coli.
XX
PN W0200044906-A2.
XX
PD 03-AUG-2000.
XX
PF 27-JAN-2000; 2000WO-US02200.
XX
PR 27-JAN-1999; 99US-0117405.
```

PA (ELIT) ELITRA PHARM INC.
XX
XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2000-514822/46.
XX N-PSDB: AAA65950.
XX
XX Novel polynucleotides and polypeptides associated with microorganism
PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
XX
PS Claim 11; Page 224-229; 316pp; English.
XX
XX AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from *Escherichia coli* which inhibit *E. coli*
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC nucleotide and protein sequences associated with *E. coli* proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing *E. coli*
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation- required gene in a microorganism, by contacting
CC a microorganism with a proliferation- required gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
XX Sequence 2383 AA;
SQ

Query Match	Similarity	5.9%	Score 231	DB 21	Length 2383
Best Local Similarity	20.6%	Pred. No. 0.00027			
Matches 161	Conservative 120	Mismatches 321	Indels 180	Gaps 30	
QY	139 KANIIVELGISVGTGDWPKRT-----VTKAAEAQFAKTDKCGTEAAKVESAK	189			
DB	1453 kagv-yglatlengsmqvtvtypnvnaaelilaasxdpriadnndltltatrade	1511			
QY	190 AVTTQKVEPKESKAVELTKEDIKTNKANKNDKVLKVEVTLSEDKRSATVELYSNLAARQ	249			
DB	1512 gnaiaetevft-----lpedvkanfltsdgg--kviltdeagr--akvllkqtkagah	1560			
QY	250 TTYTVNVNKKGTKEVAVGSLKART-----IEAADQTVVAD--PRALQPTVVDENGTENV	301			
DB	1561 tltaeatg-ykseqglvniadiadtlaqvalnvteedfiannvgmtclgtvcdngnpia	1619			
QY	302 SPEGIEFVTP-----AAEKTNAGETTLAKGTSITYKAVYKKDKGVAAESK	347			
DB	1620 n-eavftlpadvsasfljggsgaltdinglaevelfsgtksgtylvtsvnmvgysdtk	1678			
QY	348 EKKVSAEG-----AAVASISNMVTVAEQNKADFTSKQFKNKRVPEEDNNAVQVELKDQFN	402			
DB	1679 qvtliadagtaklaaistysfstvstetgaumtaasvtdangnpveg-----lknfigtav	1734			
QY	403 AVTTGKVEYES-----LNTENV-----AVDKATGKVTVLGAKAPK--VTVK	443			
DB	1735 flstsvetddgfaellvtstevglkvtasladkpteviserllinaasadvnatlsie	1794			
QY	444 DSKGALVSHVTEIEA-----FAQKAMDILEKTNVLSKRDYTDLKV	487			
DB	1795 lpegvymvaqdaavahvndqfgnpvabhpvtlfssepsqgmilsqatvstcnqyavevm	1854			
QY	488 KAPVLDDQGEFTAPVTVKVLDDKDEKLEKLEAKRYNRELVLNAAQEGANY-----	541			
DB	1855 t-----perngsyvmkasipngasl-ekglea--idekiltlasppllgvyptgat	1903			
QY	542 -TVVLTAKSGEKAKATTLALAKARGA-FSKREVR-----	574			
DB	1904 llatltsangprveqgvlnfsvcpaeglsgkvltvnssggapvvltsnkvtylvtasf	1963			
QY	575 -GLDTELDKRYTEENOKNAMTVSVLPVDANGVLTKGAFAELKVTYTNKKEKVEDATD-	631			

Db	1964	hmgvltqctvtkvrtgnstshvafiadpsltiaatntdlstlkatvedgsgnlliegltv	2023
Oy	632	-----AQTVCQNNSVITFGOGAKAGEFYKRVVLDLGGKLTITHSFVVDTPA	677
Db	2024	yfalksgsaelstltstlavddgngdtstevkqgmbsvsvtsavtaagmgqtdvdlivsgpad	2083
Oy	678	TAKGLAVEFTSTSLKEVAPNADLKAAALLNLTISVDCVPATTAKTASVVEFVSADTNY---	734
Db	2084	tsqvsiknsrsslkqdydsaeeltvlhdf---sgnplikvse---gmefvgsgtnvpyl	2136
Oy	735	-----VAENG-----TYGAGG-ATSTYVKMLTVVKKGKEKVEF---	775
Db	2137	kisaldysiningdykxatvtgggegiatlipvl-gvhqglstltlqitlrædklmsgtv	2195
Oy	776	SI 777	
Db	2196	sv 2197	
RESULT 15			
ID	AAB01835		
AC	AAB01835 standard; Protein; 1095 AA.		
XX	AAB01835;		
DT	11-SEP-2000 (first entry)		
DE	Haemophilus influenzae strain LCDC2 mature HMW1A protein, SEQ ID NO:45.		
XX			
KM	Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;		
KW	non-typable Haemophilus influenzae; NTHI; non-encapsulated;		
KM	recombinant production; Escherichia coli; antibacterial; vaccine;		
KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;		
KM	detection; diagnosis.		
XX			
OS	Haemophilus influenzae strain LCDC2.		
XX			
PN	WO200020609-A2.		
PD	13-APR-2000.		
PF	07-OCT-1999; 99WO-CA00938.		
PR	07-OCT-1998; 98US-0167568.		
PR	08-DEC-1998; 98US-0206942.		
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
PI	Loosmore SM, Yang Y, Klein MH;		
XX			
DR	WPI: 2000-303789/26.		
DR	N-PSDB: AAA52184.		
XX			
PT	Nucleic acid molecule for producing recombinant high molecular weight		
PT	proteins of Haemophilus which are used as a vaccine to provide		
PT	protection against Haemophilus induced diseases in humans -		
XX			
PS	Claim 8; Fig 22A-P; 307pp; English.		
XX			
CC	The invention relates to the recombinant production of Haemophilus		
CC	influenzae high molecular weight (HMW) proteins in Escherichia coli. The		
CC	expression construct used to effect recombinant expression comprises a		
CC	promoter functional in E. coli (e.g., the T7 promoter) operably linked		
CC	to a modified hmwABC operon from a non-typable (non-encapsulated) H.		
CC	influenzae (NTHI). Most HMW-expressing NTHI strains contain two hmw gene		
CC	clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,		
CC	hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins		
CC	and the hmwB and hmwC genes encode accessory proteins which are		
CC	responsible for post-translational processing and secretion of the HMW		
CC	proteins. The modified hmwABC operon used in the expression construct of		
CC	the invention contains an A gene modified such that it encodes only the		
CC	mature HMW. The invention also discloses hmwA genes (AAA52175-452198)		
CC	and HMW proteins (AAB01824-801849) from the non-typable H. influenzae		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 06:54:37 ; Search time 61.69 Seconds

(without alignments)
286,352 Million cell updates/sec

Title: US-09-754-947-1

Perfect score: 3865

Sequence: 1 AGKTFFPDVADHWGIDSLNY.....EFDKAVQVAVSIKKAEPATK 785

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	378.5	9.7	1232	2	US-08-682-517-15 Sequence 15, Appl
2	378.5	9.7	1232	4	US-08-682-517-9 Sequence 9, Appl
3	240.5	6.2	2353	4	US-09-377-155-33 Sequence 33, Appl
4	240.5	6.2	2353	4	US-08-913-942-4 Sequence 4, Appl
5	238	6.1	1912	1	US-08-409-995-4 Sequence 4, Appl
6	238	6.1	1912	3	US-08-685-467-4 Sequence 4, Appl
7	216.5	5.6	1529	2	US-08-728-470-10 Sequence 10, Appl
8	216.5	5.6	1529	4	US-08-719-641-10 Sequence 10, Appl
9	214.5	5.5	1600	2	US-08-617-697-10 Sequence 10, Appl
10	214	5.5	1338	2	US-08-728-470-9 Sequence 9, Appl
11	214	5.5	1338	4	US-08-719-641-9 Sequence 9, Appl
12	214	5.5	1477	1	US-08-302-832-4 Sequence 4, Appl
13	214	5.5	1477	1	US-08-302-832-4 Sequence 4, Appl
14	214	5.5	1477	2	US-08-530-198-4 Sequence 4, Appl
15	214	5.5	1477	2	US-08-469-880-4 Sequence 4, Appl
16	214	5.5	1477	2	US-08-728-470-4 Sequence 4, Appl
17	214	5.5	1477	2	US-08-617-697-4 Sequence 4, Appl
18	214	5.5	1477	4	US-08-719-641-4 Sequence 4, Appl
19	214	5.5	1599	2	US-08-617-697-9 Sequence 9, Appl
20	213.5	5.5	1183	1	US-08-447-031A-2 Sequence 2, Appl
21	212.5	5.5	1098	1	US-08-409-995-2 Sequence 2, Appl
22	212.5	5.5	1098	3	US-08-685-467-2 Sequence 2, Appl
23	212.5	5.5	1098	4	US-09-377-155-32 Sequence 32, Appl
24	212.5	5.5	1098	4	US-08-913-942-2 Sequence 2, Appl
25	211	5.4	1561	3	US-08-894-017-23 Sequence 23, Appl
26	207	5.3	1536	1	US-08-682-517-15 Sequence 2, Appl
27	207	5.3	1536	1	US-08-302-832-2 Sequence 2, Appl

28	207	5.3	1536	2	US-08-530-198-2 Sequence 2, Appl
29	207	5.3	1536	2	US-08-469-880-2 Sequence 2, Appl
30	207	5.3	1536	2	US-08-728-470-2 Sequence 2, Appl
31	207	5.3	1536	2	US-08-617-697-2 Sequence 2, Appl
32	207	5.3	1536	4	US-08-719-641-2 Sequence 2, Appl
33	200.5	5.2	1565	6	5352450-2 Patent No. 5352450
34	188	4.8	2123	4	US-08-968-685A-10 Sequence 10, Appl
35	186	4.8	679	4	US-08-913-942-15 Sequence 15, Appl
36	183.5	4.7	198	2	US-08-682-517-19 Sequence 19, Appl
37	180	4.6	1026	2	US-08-614-377A-7 Sequence 7, Appl
38	180	4.6	1026	4	US-09-142-648B-7 Sequence 7, Appl
39	176	4.5	1026	4	US-08-194-290-7 Sequence 7, Appl
40	175.5	4.5	573	4	US-08-235-836C-112 Sequence 112, App
41	174	4.5	200	2	US-08-682-517-20 Sequence 20, Appl
42	174	4.5	658	1	US-08-409-995-5 Sequence 5, Appl
43	174	4.5	658	3	US-08-685-467-5 Sequence 5, Appl
44	174	4.5	658	4	US-08-913-942-5 Sequence 5, Appl
45	172.5	4.4	1545	4	US-08-296-791-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1222 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-682-517-15

Query Match 9.7%; Score 378.5; DB 2; Length 1222;
Best Local Similarity 22.1%; Pred. No. 4.2e-18;
Matches 223; Conservative 129; Mismatches 378; Indels 279; Gaps 44;

QY	16	DSINLYEKGAVNGNDKMGFEPEKELTRAEATMMAOIILPDIKAKPFPAD-SOGQWY 74
DB	15	EAAVQALDQVIOGDNTGNPNPLNTVYRAQAIEIFKALELNGDV--EFGVKAGAWY 72
QY	75	TPFLAAVERKGVKNG-NGNFEFNGKIDRVSMASLVEAKLDTKVTGATPKFEDLETL 133
DB	73	YNSIAAVANGIEGVSATFAPNKSILTRSEAKILVEAFGLGEGAD--LSEFADASQV 129
QY	134	N-WGKEKANILVELGISVGT-GDQWEPKKTVTYKAAQFIKAT-DKQGTIEAKKVESAKA 190
DB	130	KPAKAKYLEIAVANGIEGTDANKLNPNNSITRQDFALVFRKVDKEGSTPEEAFVKA 189
QY	191	VTTQKVEVRSKRAVEKIKT-----EDIKVTN---KANKDVLV-----REVLTSED 233
DB	190	INNTVEVTEEEETVQALNFKIEGLEIKMASVKQNRKVAVLTTEAQTADDEFTVLTLD 249
QY	234	KRSATVELYSLNLAQKQTYDVANKVGKTEVAV-GSLEAKTIEAADOVTVADEPTA--LQ 289
DB	250	--GETTIGFGVAA-----VPTKVELVSSAVQKLGQEVKQAKTYVAAGSQKAGIPVT 302

QY 290 FPKDNGTEVYS-----PEGIEFVTPAA----- 313
 Db 303 FTVPGNNDDGVPTLTGALITNEEGIAITYSTRYREKGEDEVTAATGDRSKFSLGYFWG 362
 QY 314 -EKINKGEIT-----LAKGSTTVKAVYK--KDGKV-----VAESKEVKV 351
 Db 363 VDTILSVEETTGASVNNNANKTYVTKKPKRTGKPEAKNTFFNNGEYENMNVISDKVANA 422
 QY 352 SAEGAVAASISNKTVAEONKADFTSKDRONNKYYEGDNAYVOVELD-----Q 400
 Db 423 TVNGVKALDLSNGTALD--AAQITDSSKGEATFTVSGTNAVAVTEPVYDLHSTNNSTSNK 480
 QY 401 FNA-----VTTGKEVEESLUTEVAV----VDATGKVTVYLSGKA--PKVTVKSKGKALV 451
 Db 481 YKSAALQTTASKVYTPAALQAEITIELTADNA--GEVAIGATGTCNGREKVIIVKDKAGNLAK 539
 QY 452 SHTEVEE-----AFAOKAMDKILKLTN----- 474
 Db 540 NELVNAVEFNEDKDRVISTVTNAKFEVDPTPTAVYFTGDGKAKQISV--KTNDKGEATFVIGS 598
 QY 475 -----VALSRKQVTDLKVKAPVLDQYKGEFTAPV--VKVLDKDGKELKEOKLEAKVYN 526
 Db 599 DTVNDYATPIAIDINTSDAQGDIDEEEPKAVAPISYFQAPYPLDGSAIKAYK---KSDL 655
 QY 527 RELVINAAGQEGANTVYVLTAKSGEKEAKTLALTELKAPAFSKFEVBRGDTDELDDKYTE 586
 Db 656 NKAVTFKFDSEFVAPFAEELVNOSGKKVYGIS-----KKATYITLYNGADIKVDVQVIS 710
 QY 587 ENOKNAMTVSVLPVANGLVKGAEMAE-----LKVTYTT---NKEKEVDATDA 632
 Db 711 PNRSTVTYEA--TLSSGTGVTTPAKNLEWYSVDGKTTAVVIVATGIAVNTDGDYAFYAK 769
 QY 633 QVTV-----ONNSVITVQO-----GAKAGEYTK----- 655
 Db 770 EATATFTATNEVPNSYGTGATQFMTADSGSNSIMWAGKNPKYKYGVSCKITTKYFGANG 829
 QY 656 -----VTVVLDGKLITTHSEKVVVDPAFPAFGLAVEFTSTL 691
 Db 830 NEVEGEAAMEALLTOYATEGQKVTISYVNDGDIYVTFKVIS-----AVNSTEAI 878
 QY 692 KEVAN---ADLKAALLILISVDDVPATTKAKTASINVEFVSADTNYA---ENGVIYAGAKG 745
 Db 879 KVAVATTPAAATTTGALLTLTPAGSLVDLTATNTLTGISLADADLNASATVTPATVATSLKD 938
 QY 746 ATSIYVKKLVYVKDGKEQVEFDKAVO-----VAVSIEKEKPA 783
 Db 939 SANNSL-SLTLVETGANTGV-FATTVQAGTGLSLTLAGTILTVYAADAKNA 985

 RESULT 2
 US-08-682-517-9
 ; Sequence 9, Application US/08682517
 ; Patent No. 5874267
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Expression of surface layer proteins
 ; NUMBER OF SEQUENCES: 25
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/682,517
 ; FILING DATE:
 ; CLASSIFICATION:
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1252 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; IS-08-682-517-9

	Query Match	9.78;	Score 378.5;	DB 2;	Length 1252.
	Best Local Similarity	22.18;	Pred. No. 4,4e-18;		
	Matches 223;	Conservative 129;	Mismatches 378;	Indels 279;	Gaps 44;
OY	16	DSINYLVKGAVKNDKGMEPEGKELTRAAATMAQILNLPIDDKAKSPAD--SOGOMV	74		
Dd	45	EAVGLVDGVIGSDTNGNPNPLNTVTYRAQAALFFTKALELNGDV--NEKDVKAGAMV	102		
OY	75	TPTFAAVERKAGVIKG--TGNGEPNGKIDRVSMASLLVEAFKIDTKVNGTPATKFNDELT	133		
Dd	103	YNSTIAAIVANGIEGVSATFEAPNRKSLTRSEAKIITVEAFGLEGEAD---LSFADASGV	159		
OY	134	N-MGKEKANILLVELGISVGT--GDOMEKKIKVTAEAAOPIAKI--DKQFTGAKEAVESAKA	190		
Dd	160	KPAKKKLETLAVANGIEGEDDANKLNPNNSITRQDRLAFKRKYDVKVEGETPEEAFAVA	219		
OY	191	VTTOKVEPKSKAVEKLTK-----EDIKVYN--KANNDRVLV-----KEYTLSED	233		
Dd	220	INNTVEVTEPEEEETVNOALNFRIEGLETKNASVKOTNNKKVVYLTTETAQTAKEVYLTLD	279		
OY	234	KRSATVLYENLNAKQTYIVDAVKVKTERTAV--GSLEAKTIEMADQTVVADEPTA---LQ	289		
Dd	280	--GETTGGFGGVA-----VPTKELVSAVOGKIGQEVKKVOAKTVYAEGOSKAGIPVT	332		
OY	290	FIVKENGTEFVVS-----PEGIEFTVPAA-----	313		
Dd	333	FIVPAGNNDDCVPTLGEALTNNEGIALITYSYTRYRKGTDEVTARAYATGDRSKFSLOIYFWG	392		
OY	314	-EKINKAGETT---LAKGSTTTKAVNY--KDGKY-----VAESEKENV	351		
Dd	393	VDTILSVEEETTGASVNNQANKRYTKYTNPRTGKREAPNKTEPVGEVENMNTSDKVANA	452		
OY	352	SABGAAVASISMTVAEOKKADPTSFDFCONKKVYEGDAAVQVELKD-----Q	400		
Dd	453	TVNGVALDLQSLNTALD--AAQTIDSKEAFETYSGTNAAVTPPVYVDIHSTNNSTSARK	510		
OY	401	FNA---VTTGKEVESLNTAV----VDKATGRVTVLSAGKA--PYRFTVYKDSKGAUV	451		
Dd	511	YSASALQTTASKVTFEALQAEYTIELTRADNA--GEVAALGATNGREXYVIYKDAQNLAK	569		
OY	452	SHIVEE-----AAQKAMDIKILEKN-----	474		
Dd	570	NELVNAVEFDKDRVISTVTAKEVYDDEDPNAVYPTGKAOAISV-KTNDKGEATFYIGS	628		
OY	475	-----VALSTRDYDLVKAPVLDJOYEFTAFTV--VKULDKRGKELKROKLEAKVYN	526		
Dd	629	DIVNDYAPFIAMIDINTSPAQKGDLBEBEPRAVARAISIFQAPYLDGSAIKRYK---KSDL	685		
OY	527	RELVLNAAQOEAGNTVVVLTAAGSEKAKATLALELKPAGFSEFVEYGLDELDKYYTE	586		
Dd	686	NAAYVRFDGSEFAVFAELVNOSGKKVYGTSI---KRAYTIIYNGANDIKVDNQVIS	740		
OY	587	ENOKNAATVSYVLPPVANGJVLKGAFAA-----LKTTYT---NREGREVADTDA	632		
Dd	741	PKRSTVITYEA-TLSSGTGITPRANLEVTISVDGKTAAVKIAGIAGVINDGCKQYAFETAK	799		
OY	633	QVTV-----ONNSYITVGQ-----GAKAGETRYK-----	655		
Dd	800	EATATFATINEVPNSYTGATQFMADGSGSNSIMWPGAKRPKVPYAGVSGKITTYFPBAG	859		
OY	656	-----VTYVLDGKLITTHSEFKVVDTPAPKAGLAVEFTYSTL	691		
Dd	860	NEVEGEAAWEALLTOYATEGOQVITISYVNDGDTV---FKVIS-----AVNSSTEAI	908		
OY	692	KEVAPN---ADLKAAALLIISVGDVPATFTAKATSNSEFVASADPNVYA---ENGTYGAKG	745		
Dd	909	KPVAPTPTAAATTGALLITLPAGGLVLDLTATNTLGISLADADLNASTATTVADPATSLDK	968		
OY	746	ATSIFYXNVLVKDKQKVEEDRAVO-----VAVSIRKEAKPA	783		
Dd	969	SANNST-SLIVETGANDGV-FATTVAGTIGTSSITACTCTVTVTAADAKNA	1015		


```

Db 79 EKEVTENSMNGIEDNKGVLKAGAITLKAGD-NLKIKONTDESTINASSFTYSLKDLTDL 137
OY 78 IA-AVERAGV-----IKGTGNFE-----PNGKI-----DRVSNASLLEAVK 114
Db 138 TSVATEKTSFPGANDKDVITSDANGKLAKTGNVHNLNGLDSTLPDAVITMTGVLSSSF 197
OY 115 LDTKVNKGTPTATKFDLETLNNGKEKANILVELGISVGTGDOMEKPKTYTKAEAOFIKT 174
Db 198 TPNDVEKTRAAVTYKDVLAAGNNIKGAKT-----AGNVESVDLVSAVNNVEFTIGD 248
OY 175 DKQFTEAKVESAKAVTTQKVEVESKAVEKTEKEDIKVTNKNAND-KVLVEKTVLS 231
Db 249 KNTLDVVLTAENKGT-----TEVFTPKTSVIKKEKDKLFTGKNNDTKVTSTATDN 303
OY 232 EDKRSATVELYSNLAOKOTYVDNKGK-----KTEVA-----VGSLEAKTIEADOT 279
Db 304 TDEGNGLVTAKAVIDA-----VKNAGMRYKTTTANGONGDFATVAGTNTVFESGDT 356
OY 280 VVADPTALOFTVKDENGTEVVS-----EGIEFVTPAAEKINAKGEITLAKGSTTVK 333
Db 357 TAS-----VTKDTNGNGITVKYDAKVGDLKF--DSDKI-----VADTTALTVT 399
OY 334 AVYKKKGKVAEESKE---VKVSAEGAAYASISN--WTVAEQNKADFTSKDFQNNKRYEG 388
Db 400 G-----GKVAELIAKEDDKKILVNAAGDLVTALGNLSMKAKAEADTDGALGISKDOEVKAG 454
OY 369 DNAYVOVELKDOPNAVTTGKVEYESLNTPEVAVVDKATGKTV-----LSAGKAPVYTVK 443
Db 455 ETV-----TFKAGKNLKVKODGANFTYSLODALITGLTTLGGTTNGGDAKTVINK 506
OY 444 D-----SKGALVSHVTEIEFAOKA-MKDIKLEKTVALSTKDVTDLKVKAPVLDQY 495
Db 507 DDLTTTPAGNGTGTGTNTISVTIKDKIKAGNKAI---TNVAGLIRAYD-----551
OY 496 GKEFTAPVTVKVLKDKGKELKEQ-----KLEAKYVNR-LVLAAGQEAQNTYV--543
Db 552 -----ANFDVLNNSATDLNRHVEDAYKGLNLNEKNANKOPLVTDSTATYVGDRLKLG 604
OY 544 -VLTAKSGEKEA--KATLALTELKAPGAFSKFEYRGIDTELDKRYTEENOKNAMTVSLPV 600
Db 605 WYVSTKNGTKRESNOVKQKODEVLFTGA-----GAATYTSK--SENGHNTTVSAET 654
OY 601 DANGLVKAEGAELKVTYTNKEGEVDAITDAQVTVQNNNSVITVGGACAAGEYTKTVVL 660
Db 655 KADCGLEKGDITIKLVNDON-----TDNLVTGNN-----665
OY 661 DGLLTHSFKVVDPATPAKGLAVEFTSLKEVAPN-ADLKAALINIISVDGVATTAK 719
Db 686 -GTAVTKGGFEYVTKGAT-----DADGKYTVKDATANDADKKVATYKDV-----ATAIN 734
OY 720 ATASNVEFVSADTNVAENGTVGAK-----GATSIYV--KNLTYYKDGKEQKVEVDKAV 771
Db 735 SAATEFKTEULTISIDENDTNGKDDALKAGDTLTFKAGKNLKVXRDKG--NITEDLAK 792
OY 772 QVAVSIKEAK 781
Db 793 N--LEVKTAK 800

```

```

OY 5
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA

```

```

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-4

```

```

Query Match      6.1%; Score 238; DB 1; Length 1912;
Best Local Similarity 22.8%; Pred. No. 56-08;
Matches 194; Conservative 101; Mismatches 339; Indels 216; Gaps 42;

```

```

OY 12 HMGIDSINTLVKEGAVK-----GNDKGMFEPKGELETRAEAAVT--MMAQILNLPIDKD 61
Db 86 NMGI-----YFDNKGVLKAGAITLKAGDNLKXQXTDEXTNASSFTYSLKDLTDLVSAT 141
OY 62 AKPSFADSGOWTPTPIAAVERKGVK--GTGNG-PEPNG---KIDRYSMASLLEAVK 114
Db 142 EKLSFPGANDK-----VDITSDANGKLAKTGNVHNLNGLDSTLPDAVNTGVLSSSF 196
OY 115 LDTKVNKGTPTATKFDLETLNMGKEKANILVELGISVGTGDOMEKPKTYTKAEAOFIKT 174
Db 197 TPNDVEKTRAAVTYKDVLAAGNNIKGAKT-----AGNVESVDLVSAVNNVEFTIGD 247
OY 175 DKQFTEAKVESAKAVTTQKVEVESKAVEKTEKEDIKVTNKNAND-KVLVEKTVLS 231
Db 248 KNTLDVVLTAENKGT-----TEVFTPKTSVIKKEKDKLFTGKNNDTKVTSTATDN 302
OY 232 EDKRSATVELYSNLAOKOTYVDNKGK-----KTEVA-----VGSLEAKTIEADOT 279
Db 303 TDEGNGLVTAKAVIDA-----VKNAGMRYKTTTANGONGDFATVAGTNTVFESGDT 355
OY 280 VVADPTALOFTVKDENGTEVVS-----EGIEFVTPAAEKINAKGEITLAKGSTTVK 333
Db 356 TAS-----VTKDTNGNGITVKYDAKVGDLKF--DSDKI-----VADTTALTVT 398
OY 334 AVYKKKGKVAEESKE---VKVSAEGAAYASISN--WTVAEQNKADFTSKDFQNNKRYEG 388
Db 399 G-----GKVAELIAKEDDKKILVNAAGDLVTALGNLSMKAKAEADTDGALGISKDOEVKAG 453
OY 444 D-----SKGALVSHVTEIEFAOKA-MKDIKLEKTVALSTKDVTDLKVKAPVLDQY 495
Db 506 DGLTTPAGNGTGTGTNTISVTIKDKIKAGNKAI---TNVAGLIRAYD-----550
OY 496 GKEFTAPVTVKVLKDKGKELKEQ-----KLEAKYVNR-LVLAAGQEAQNTYV--543
Db 551 -----ANFDVLNNSATDLNRHVEDAYKGLNLNEKNANKOPLVTDSTATYVGDRLKLG 603
OY 544 -VLTAKSGEKEA--KATLALTELKAPGAFSKFEYRGIDTELDKRYTEENOKNAMTVSLPV 600

```

Db 604 WVSSTKNGTKEESNOVKQADAEVLFTGA-----GAAVTISK-----SENGKHTITVSAET 653
Qy 601 DANGLVJLKAFAELKVTYTTKKEGEVDATDAQYTVONNSYITVGGAKAGETVKVTVL 660
Db 654 KADCGLEKDGDTIKLKVDNQN-----TDNVLTVGN-----684
Qy 661 DGKLTTHSFVVDYAPLAKGLAVEFTSTLSKEVAPN-ADLKAALLNLSVDPVATTAK 719
Db 685 -GTAFTKGFETVKTGAT-----DADRGKVTVKDATANDADKKVATVKDY-----ATAIN 733
Qy 720 ATASNVEFVSADTVNVAENGTVGAK-----GASITVY--KNLTVKDGKQKVEFDKAV 771
Db 734 SAAFTVKTENLTSTIDEDNPTDNGKDALKAGDTLTFKAGKLNKVRDGGK--NITFDLAK 791
Qy 772 QVAVSIREAK 781
Db 792 N--LEVKTAK 799

RESULT 6
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embardadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/REF/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELETYPE: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-685-467-4

Query Match 6.18; Score 238; DB 3; Length 1912;
Best Local Similarity 22.88; Pred. No. 5e-08;
Matches 194; Conservative 101; Mismatches 339; Indels 216; Gaps 42;
Qy 12 HMGIDSYNLYVEKGVK-----GNDKGMFEPKELTREAAT--MMAQILNLPIDKD 61
Db 86 NMGI-----YFDNKGVLKAGATTLKAGDNLKKKQATDEXTNASSFTYLSLKKLDTLTSVAT 141

Qy 62 AKPSFADSGQWYTPPIAAVEKAGVYK--GTGNG-FEEDNG-----KIDRVSNASLVEAYK 114
Db 142 EKLSPGANGDK-----VDITSDANGKLAKTGNVHLNGLDSTLPDAVTMTGVLSSSSF 196
Qy 115 LDTKYNTPATKRFKDLFTLNKGERKANILVELGISVGTGDMEPKYTKTKEAEOFIAKT 174
Db 197 TPNDVEKTRAAVTVDVLAAGNNIKGAKT-----AGGVNESVDLSAANNVEFTIGD 247
Qy 175 DKQFTEAKAVESAKAVYTKQVEYKFSKAVERLTKEDIKY--TKRANN--TVLVEVTL 231
Db 248 KNTLDVULITAKENKT-----TEVKFTPTSVIEKDKLTGKENNDTNKVTSTATAIN 302
Qy 232 EDKRSATVELYSNLAQOTYTVDNKVG--KTEVA-----VGSLEAKTIEMADOT 279
Db 303 TDEGNGLVTAKAVIDA-----VNKAGMRYKTTANGONGDFATVAGSTNTFESDGT 355
Qy 280 VVADEPTALQFTVDENGTEVYSP-----EGIEFVTPAAEKINAKGKITLAKGSTIVK 333
Db 356 TAS-----VTKDTNGNGITVYKDAKVGDLKF--DSDKI-----VADTTALTVT 398
Qy 334 AVYKDKGVVAESKE---VKYSAEGAASISN--WVAEQNKADFTSKDKQNNKYEG 388
Db 399 G-----GKVAELAKEDDKKLLVNAQDLVTALGNLSMKAKAEADTDGALGSKDQEVAKG 453
Qy 389 DNAYVQVELKDQFNAVTTGKVEYESLNTFVAVDKATGKTVY----LSAKRAPHYKTVK 443
Db 454 ETV-----TFKAGKMLKVKQDGNFTYSIQDALGTLSITLGTTNGSNDAKTVINK 505
Qy 444 D-----SKKALVSHTEVEIAFAQKA--MKDIKEKTNVALSTVDVLDVYKAPVLDQY 495
Db 506 DGLTTPAGNGTGTNTISYTKDGIKAGNKAI-----TNVAGSLRAYD-----550
Qy 496 GKEFTAPVTVKVLDDKQKELKEQ-----KLEAKTVNRE-LVLAAGQEGAGNYTV--543
Db 551 -----ANPDVLNNSATDLNRHVEDAKGLLNLEKANNQPLVTDSTAFVGLRLKG 603
Qy 544 -VLAKSGEKA--KATLLELKAPGAFSKFEVRGLDTLKYTEENQNNMVTSLPV 600
Db 604 WVSSTKNGTKEESNOVKQADAEVLFTGA-----GAAVTISK-----SENGKHTITVSAET 653
Qy 601 DANGLVJLKAFAELKVTYTTKKEGEVDATDAQYTVONNSYITVGGAKAGETVKVTVL 660
Db 654 KADCGLEKDGDTIKLKVDNQN-----TDNVLTVGN-----684
Qy 661 DGKLTTHSFVVDYAPLAKGLAVEFTSTLSKEVAPN-ADLKAALLNLSVDPVATTAK 719
Db 685 -GTAFTKGFETVKTGAT-----DADRGKVTVKDATANDADKKVATVKDY-----ATAIN 733
Qy 720 ATASNVEFVSADTVNVAENGTVGAK-----GASITVY--KNLTVKDGKQKVEFDKAV 771
Db 734 SAAFTVKTENLTSTIDEDNPTDNGKDALKAGDTLTFKAGKLNKVRDGGK--NITFDLAK 791
Qy 772 QVAVSIREAK 781
Db 792 N--LEVKTAK 799

RESULT 7
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlaire, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 5.6%; Score 216.5; DB 2; Length 1529;
Best Local Similarity 22.5%; Pred. No. 1.1e-06;
Matches 174; Conservative 109; Mismatches 328; Indels 161; Gaps 37;

QY 89 CTGCGFENGCKIDRYSMSLLEAVEAKIDTKYNGRPATKFKDELTLNMGKREKANILYELGI 148
DB 769 GENSSSTITGNINITNKANVTLOA---DTSNSNTGLKK---RTITLG---NISVEGNL 817
QY 149 SVGTGDEWEPKRYTKAEAAOFIAKTDKQFSTEAKYVESAKAVTQK---VEYKESKAVEK 206
DB 818 SL-TGANANINIGNLSIAE-----DSTFKGEASDNINITGTFNNGTANININIKQYVK 868
QY 207 LTKEDIK-----VTNKANDKVLV-----EVTLSDEKRSATVELYSNLAKO-T 250
DB 869 L-QGDINNKGGLNITTNASGQTKIINGNITNEKGLNINKAKADAEIOIGINISQKEGN 927
QY 251 YTVQDNKVG-----KTEVAVGSLEAKTIMADQTVVADE-PTALOFTVKDENGTEVVS 302
DB 928 LTISSDKYNITNOITIKAGVEGGRSDSEAEANLITOTKELKLAGDLNISFNKAELTA 987
QY 303 PEGIEFTVPAEKINAKG-EITLAKGTSTYKAVYKKDKYVAESKEVKSAGAASVSI 361
DB 988 KNGSOLTTIGNASGNADAKKVTFRDKVDSKIST---DGHVYTLNSEVKS-NGSSNAGN 1042
QY 362 SNMYVAEONKADFTSKDKONKKYVEGDNAYVQVELKD--QFNATYTKGVEESLN---- 415
DB 1043 DNSTGULTISADVYVNNVNTSHKTIINISAAAGNVTTKEGTINA-TTGSVEVTAONGTIL 1101
QY 416 -----TEVAVVDKATGKVTYLSAGKAPVVTYKDSKGLVSHTEI 457
DB 1102 GNITSQNTVATENTVLTENAVINATSGTYNI-STKTDIGIGIESTSGNANNITASGNT 1160
QY 458 EAFACKAMKDIKLEKTNVALSTKDVTDLKVAPVLDQYKGEFTAPVYKVLDDCKEELKE 517
DB 1161 LKVSNTIGQDVTATADACALTTTASSTISATG-----NANITTKTGIDNG----- 1206
QY 518 OKLEAKYVNBELVLAAGOEAGN---TVVLAKSGEKEAKATLLELKAPAFSKFEVR 574

DB 1207 -KVSSSGSVTLVATGATLAVGNISGNTVITPADSG-----KLSTVGSTINGTNS----- 1256
QY 575 GLDFELDKYVTEENOK-----NAMTVSLVPVANGVLKGAFAELK---VTTNKEG 624
DB 1257 -----VTTSSQSGDIEGTISGNTVNV-TASGDLTIGNSAKVEKKNAATLTAESG 1306
QY 625 KEVDATDAQTVQVON-NSVITVGGAGAKAGETRYKVTVLD--GKLITTHSEKVVDTAPPAKG 681
DB 1307 KLTTQGTSSITSSNGQTTLTFKDSIAGNINANAVNTLNTGTGLTTGDSK----- 1356
QY 682 LAVETSTSLKEVAPNADLKAA-----LNLISVDGVPATTAKTASVNEFVSND----- 731
DB 1357 --INATSGTLTINAKDALDAAASGDRTVNATNASGSGNVTAK-TSSSVN-ITGDLMTI 1412
QY 732 --TNVVAENG--TYGAKGATSIYKNLVTVKDGKQKVEFPKAVOVAISKE 779
DB 1413 NGLNITISEGNRYVLRG-KEIDVK---YIOPGASVEVEIAKRVLEKVD 1460

RESULT 8
US-08-719-641-10
Sequence 10, Application US/08719641
Patent No. 6218141
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/719,641
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-10

Query Match 5.6%; Score 216.5; DB 4; Length 1529;
Best Local Similarity 22.5%; Pred. No. 1.1e-06;

Oy 628 DATDAQTVON-NSVITVGAGAKAGETVYKVYLD--GKLITTHSFKVVDPAPTAKGLAV 684
 Db 1380 TQTSSSITSSNGOTTLTAKDSSSAGNINAAVNTLTGTTLTTGDSK-----I 1427
 Oy 685 EFTSTSLKEVAPNADLKAA-----LNLILSVDPATPTAKATASNVEFSAD-----T 732
 Db 1428 NATSGTLTINAKDLGAASGDRTVVATNATNASSGSGNVTAK-TSSSVN-ITGDLNTINGL 1485
 Oy 733 NVVAENG--TVGAKGATSIYVKNLTVKDKGEQKVEFDKAVQVAVSIKE 779
 Db 1486 NIISNGRNTVRLRG-KEIDVK---YIOPGVAIVEVEIAKRVLEKVD 1530

RESULT 10
 US-08-728-470-9
 ; Sequence 9, Application US/08728470
 ; Patent No. 5928651
 ; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,470
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/302,832
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US PCT/US93/02166
 ; FILING DATE: 16-MAR-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 9205704.1
 ; FILING DATE: 16-MAR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-633
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1338 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-728-470-9

Query Match 5.5%; Score 214; DB 2; Length 1338;

Best Local Similarity 20.4%; Pred. No. 1.4e-06; Matches 194; Conservative 139; Mismatches 361; Indels 258; Gaps 46;

Oy 5 PPVDPADHWGIDISINYLVEKGA---VKGNDKGMEPEPKELTRAE-----AATMMAO- 52
 Db 399 YRDGGRYVWNTTILN--VTSGSKFNLSIDSTGSGSTGFSINNAELNGITTFKATFNINAG 456

Oy 53 -LNLPIIDKADPSFADSGOGWTPPTAAVEKAGVINGKNGFEPNGKIDREVSNASLVE 111
 Db 457 STANPSIKASIMPRKSNAN---YALFNEDISVSG---GGSVNFKLNASSNIQTPGVIIK 510
 Oy 112 AYK-----LDTRKNGTPATKFKDLETLMWGKEKANIYELGISVGTGOMPEKTYT 163
 Db 511 SONFNVSOGSTLNLKAESETETAFESIENDLNLNATGCGITIRO--VEGTDSR-----VN 562
 Oy 164 KAAEAQFIKTDQGFETAAKAVESAKAVTQKVEVFSKAEKLEKEDIKYTNKANNDKV 223
 Db 563 KGVA-----AKKNITFFKGSNTITFGSOKAT---EIKGVNTINKNTNATBRANPAENKSP 614
 Oy 224 -----LVKEVYLSDEKRSATVELYSNLAQOTYVDV-----NKVGTEVAV 265
 Db 615 LNIAGNVINNGNLPTAGSIINIAGNLVFSKGANLQATINTYFVNAAGSPDNNGASNISAR 674
 Oy 266 GSLAKTI-EMADOTVVADEPTALQFYVK-----DEN-----GTEVVSPE 304
 Db 675 GGAKFKDINNTSSLNITNSDTTYRTTIKGNISKSGDLNIDKSDAEIQIGNISQKE 734
 Oy 305 GIEFVTPAEKIKNKGETTLAKGT---STVKAIVYKKDKGVVAESKEVKSAGGAAYA 360
 Db 735 G-NLITSSDKVNITNOITTIKAGVGGSDSEA---ENANLTIOTKELKLAGD-----LN 785
 Oy 361 ISNMTVAEONKRAFTSK-----DEKONKV-YEGDNAVVOVEL 397
 Db 786 ISGF-----NKAEITAKNGSDLTIGNASGNADAKVTFPDKVSKISTDGHNTLNSEV 840
 Oy 398 KDOFNNAVITTK-----VEYSINTEVAVVDKATGKVVLGSKAPV 438
 Db 841 KTSNGSSNAGDNDSTGLTISAKDVTVNNNTVSHKTINISAAAGVTTKREGTTINTTGSV 900
 Oy 439 KVTYKDSKKA-LVSHFVEIEA---FAOKAMDKLEKTNVALSTKDTV-----D 484
 Db 901 EYLNQNGTIKGNITISQNTVATATELVTENAVINATSGVINISPKTDIGKIESTSGN 960
 Oy 485 LKVA-----PYLDQKKEFT-----APVYKYLDKQKELKE 517
 Db 961 VNITASGNTLKVNSITGODVVTADAGALTATAGSTISATGNANITTKTDING----- 1015
 Oy 518 OKLEAKYVNRRLVLAQAQEGANY---TVVLTAKSGEKEAKATLLEKAPGAFKFEVR 574
 Db 1016 -KVSSSGSVTLVATGATLANISGNTVTTIADSG---KLSTVSGTNGTNS----- 1065
 Oy 575 GLDELDKYYTEENOK-----NAMTVSLVPDANGVLKGAEAELK---VTTNKEG 624
 Db 1066 -----VTTSSQSGDIEGTISGNTVNV--TASGDLTINSKAVEKKNAAATLTAESG 1115
 Oy 625 KEVDATDAQTVON-NSVITVGAGAKAGETVYKVYLD--GKLITTHSFKVVDPAPTAKG 681
 Db 1116 KLTTQTGSSITSSNGOTTLTAKDSSSAGNINAAVNTLTGTTLTTGDSK----- 1165
 Oy 682 LAVETSTSLKEVAPNADLKAA-----LNLILSVDPATPTAKATASNVEFSAD----- 731
 Db 1166 --INATSGTLTINAKDLGAASGDRTVVATNATNASSGSGNVTAK-TSSSVN-ITGDLNTI 1221
 Oy 732 --TNVVAENG--TVGAKGATSIYVKNLTVKDKGEQKVEFDKAVQVAVSIKE 779
 Db 1222 NGLNISNGRNTVRLRG-KEIDVK---YIOPGVAIVEVEIAKRVLEKVD 1269

RESULT 11

US-08-719-641-9
 ; Sequence 9, Application US/08719641
 ; Patent No. 6218141

; GENERAL INFORMATION:
 ; APPLICANT: Barenkamp, Stephen J
 ; TITLE OF INVENTION: High Molecular Weight Surface Proteins
 ; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd.
 ; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

QY 384 KYEGDNAVQVELKDOFNAVTTGKVEYESLNFVAVDVKATGKVTVLSAGKAPVTVK 443
; Sequence 4, Application us/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: us/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERSSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-530-198-4

Query Match 5.5%; Score 214; DB 2; Length 1477;
Best Local Similarity 19.9%; Pred. No. 1,6e-06;
Matches 167; Conservative 122; Mismatches 311; Indels 238; Gaps 39;
QY 2 GKTFPPVADHMG-----IDSTNYLVEKA-----VGNCKMGMEPEHQLRAAAT 48
Db 748 GSVFFDIYANHSGGAEKLMKSEIN--ISNGANFTLNHVGDD--ARKINDDL-----T 797
QY 49 MAAQILPLPIDKDAKPSFADSGOWPPTFAAVEKAGVIGTGNGCFPNCKIDRVSMAAS- 107
Db 798 INATNSFSL-RQTKDFFYGYAR-----NAINSTYINISLIGGNVTLGGQNSSSTTCN 850
QY 108 -LLVEAYKLDTRYKNGTPATKFEKDLTLNMGKEKANILVELGISVGTGDWPKKTVTKAE 166
Db 851 ITIEKAAVNTLEANNAPNOQINIRDRVILKG---SLVNGSLSL--TGEMADIKGNLTISE 905
QY 167 AAOFIATDKQFGEAAKVESAKAVTTQKVEYFSKAVEKL---TREDIHYTNKANDK 222
Db 906 SATFKGTRDTLNTIGNFTNNGTA-----EINTQGVAKLGAVTNDGDLNITTHAKRNQ 959
QY 223 -----VLVKE--VTLSDEKRSATVELYSNLAQK--TYTVDVAKVGKT--EVAVGSLEAK 271
Db 960 RSLTIGDITINKKSLNITDSNNDAEIQIGNISQKBEGLNLTSSPKINITKIDITL-----KK 1015
QY 272 TIEMADQTVVADEPTALQFTVVKDENGTEVVSPECIEFVTPAEKINKAGETITLAKGST 331
Db 1016 GIDGEDSSDATSANANLTJTKELKLTEDLISGFN-----KAEITAKDGRDLT 1064
QY 332 VKAVYKKDKGVNAESKEY-----KVSAGAVALASISMTVAEQKADFTSKDFQNN 383
Db 1065 IG--NSNDGNSGAERAKVTENNPKDSKISADGHVNTLNSKVTSSSGRRSSND----- 1117
QY 384 KYEGDNAVQVELKDOFNATVGTGKVEYESLNFVAVDVKATGKVTVLSAGKAPVTVK 443
Db 1118 -----NDGILITIAKNEVVKNDITSLKVTNITASKEVTTAGS-----TIN 1158
QY 444 DSKGALVSHTEIEAFQAQAMKDIKLEKTINVALSTKDVTDLKYKAPVLDYQKEFPAPV 503
Db 1159 ATNGKA-----SITTKTGDISG-----TISGNTVSVA 1186
QY 504 TVKVLKDKGKELKQKLEAKYVNRRELVNMAOGAGVY---TVVLTAKSGKEKAKATLAL 560
Db 1187 TVDLTTKSGKITEAKSGEAN-----VTSATGTTIGTISGNTVAVTNAG----- 1231
QY 561 ELKAPGAFSFEVAGLDTLDELKDYTEENOKNAMTVSVLPVDANGVLVGAFAELKVTY- 619
Db 1232 -----LTV-----GNGAEINATGGAATLTAATG 1253
QY 620 ---TNKEGKEVDATDAQVTV--QNSSVITVGGAKAGETKYVTVLDGKLITTHSEKVD 674
Db 1254 NLTTEAGSSITSTKGVDLLAONGSI-----AGSINAAVNTLN-----TTG 1295
QY 675 TAPTAAGLAEFTSTSL-----KEVAPNADLK--AALLNITSVDGVPATTAATASVNEFV 728
Db 1296 TLTTVAGSDIKATSGTIVINAKDKLNGDASGSTEVNAAVNASGSGSVTA-ATSSSVN-I 1353
QY 729 SADTNVAENGTVGAKGATSIVYKNLTVVKGKQKVEFPDKAVOVAV-SIKEAPATK 785
Db 1354 TGDINTVNGLIITISKDRNTVRLR-----GKEIEVKY---IQPGVASYVEVIEAKR 1401
RESULT 15
US-08-469-880-4
; Sequence 4, Application us/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: OF NO. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
US-08-469-880-4

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 06:57:02 ; Search time 66.49 Seconds

(without alignments)
899.339 Million cell updates/sec

Title: US-09-754-947-1

Perfect score: 3885

Sequence: 1 AGKTFPPVADHMGIDISINY.....EFDKAVOVAVSIEKAPATK 785

Scoring table: BLOSUM62

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : PIR.68:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3847	99.0	814	2 I40048	S-layer protein pr
2	760	19.6	874	2 JC4930	S-layer protein pr
3	445	11.5	1099	2 T14850	S-layer protein pr
4	402.5	10.4	1176	2 A33856	surface-layer 125k
5	279.5	7.2	404	2 F59097	hypothetical prote
6	268	6.9	1036	2 T30311	S-layer protein -
7	259	6.7	1109	2 A56143	surface-array prote
8	255.5	6.6	762	2 A34355	cell surface prote
9	248.5	6.4	1228	2 I40468	surface layer prote
10	245.5	6.3	1832	2 S33441	EF protein - strep
11	241	6.2	1128	2 T30296	R27-2 protein - Tr
12	240	6.2	3468	2 T34418	hypothetical prote
13	237.5	6.1	752	2 B59102	hypothetical prote
14	236	6.1	745	2 B33856	hypothetical 80K p
15	234.5	6.0	1231	1 A48490	endo-1,4-beta-xyla
16	234.5	6.0	6642	2 S72957	protein UNC-89 - C
17	232.5	6.0	1087	1 S41797	cellulose 1,4-beta
18	232.5	6.0	1829	2 A25833	hypothetical prote
19	231.5	6.0	1116	2 A35129	surface layer prote
20	231	5.9	2383	2 D64962	probable membrane
21	228.5	5.9	2055	2 T31110	extracellular matr
22	220.5	5.7	1148	2 S72635	exo-poly-alpha-gal
23	218.5	5.6	1302	1 JC6009	surface-located me
24	217	5.6	1324	2 T18265	endo-1,3(4)-beta-g
25	216.5	5.6	1417	2 D85521	probable adhesin e
26	215.5	5.5	1234	2 S72640	endo-1,4-beta-xyla
27	214	5.5	1477	2 B43855	high-molecular-wet
28	213.5	5.5	1185	2 A42404	collagen adhesin
29	212.5	5.5	1556	2 A60988	saliva-interactin

30	212	5.5	2139	2 T18296	myosin heavy chain
31	211.5	5.4	1365	2 T30822	Impl protein - Myc
32	211	5.4	1561	1 S06839	surface antigen sp
33	210	5.4	2269	2 T28677	rhodury protein -
34	210	5.4	1315	2 T28679	fibrinogen-binding
35	207	5.3	1536	2 A43855	high-molecular-wet
36	205	5.3	837	2 JN0292	antigen 332 - mala
37	205	5.3	2660	2 E85822	probable invasin z
38	204	5.3	1084	2 A28555	middle wall protei
39	203.5	5.2	2032	2 T39917	hypothetical prote
40	203	5.2	1361	2 T03415	S-layer protein -
41	201	5.2	1565	2 S04729	surface antigen pa
42	199.5	5.1	631	2 JC6031	stafoldin dockeri
43	199.5	5.1	1983	2 G86643	hypothetical prote
44	199	5.1	941	2 S29043	cellulase (EC 3.2.
45	196	5.0	1503	2 T01098	chloroplast outer

ALIGNMENTS

RESULT 1	140048	S-layer protein precursor - Bacillus anthracis	
C:Species: Bacillus anthracis			
C:Date: 02-Aug-1996	#sequence: revision 02-Aug-1996	#text: change 21-Jul-2000	
C:Accession: I40048;	S51695		
R:Etienne-Tourelle, I.; Siraad, J.C.; Duflot, E.; Mock, M.; Fouet, A.			
J. Bacteriol. 177, 614-620, 1995			
A:Title: Characterization of the Bacillus anthracis S-layer: cloning and sequencing o			
A:Reference number: I40048; MUID:95138020			
A:Accession: I40048			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-814 <RES>			
A:Cross-references: EMBL:Z36946; NID:q1359499; PIDN:CAA85408.1; PID:q634013			
C:Genetics:			
A:Gene: sap			
C:Superfamily: S-layer repeat homology			
F:34-88/Domain: S-layer repeat homology <SLR1>			
F:95-147/Domain: S-layer repeat homology <SLR2>			
F:156-209/Domain: S-layer repeat homology <SLR3>			
Query Match	99.0%; Score 3847; DB 2; Length 814;		
Best Local Similarity	99.0%; Pred. No. 3.4e-154;		
Matches 777; Conservative 4; Mismatches 4; Indels 0; Gaps 0;			
QY 1	AGKTFPPVADHMGIDISINYLVKGAAGVKGNGEPNCKIDRVASASLVVAYKLDFTVN 120		
DB 30	AGKTFPPVADHMGIDISINYLVKGAAGVKGNGEPNCKIDRVASASLVVAYKLDFTVN 120		
QY 61	DAKSPFADSOQOWTTPFLAAVEKAGVKGNGEPNCKIDRVASASLVVAYKLDFTVN 120		
DB 90	DAKSPFADSOQOWTTPFLAAVEKAGVKGNGEPNCKIDRVASASLVVAYKLDFTVN 149		
QY 121	GTPATKFKDLETLMMGKEKANIIVELGISVGTGQWBEKKITVTAEAQFIAKTDKQGT 180		
DB 150	GTPATKFKDLETLMMGKEKANIIVELGISVGTGQWBEKKITVTAEAQFIAKTDKQGT 209		
QY 181	EAKVESAKAVTTQKVEKFAVEKLEKEDIKVTNNKANDKVLVKEVTLSEDKRSATVE 240		
DB 210	EAKVESAKAVTTQKVEKFAVEKLEKEDIKVTNNKANDKVLVKEVTLSEDKRSATVE 269		
QY 241	LYSNLAKQTYTVVNVKGEVAVGSLEAKTTEMAQDTVADEPTALQFTVKDENGTEV 300		
DB 270	LYSNLAKQTYTVVNVKGEVAVGSLEAKTTEMAQDTVADEPTALQFTVKDENGTEV 329		
QY 301	VSPGIEFTVPAAEKINAKGEITLAKGSTTVKAVYKDKGVVAESKEVKYSAGAAVAS 360		
DB 330	VSPGIEFTVPAAEKINAKGEITLAKGSTTVKAVYKDKGVVAESKEVKYSAGAAVAS 389		
QY 361	ISNMTVAEDONKADFTSKDFKONKNKVVYEGDNNAVVOVELKDQFNAVTTGKVEYSINTEYAV 420		

```

|||||
Db 390 ISNMVVAEONKADFTSKDFKONNNKYEGSNATVVOVELKQFNPANVTGKVEYESLNTEVAV 449
Oy 421 VDKATGKVTLSAGKAPKVTYVDSKGAALVSHVEIEAFPAOKAMKDILEKTVALSTR 480
Db 450 VDKATGKVTLSAGKAPKVTYVDSKGAALVSHVEIEAFPAOKAMKDILEKTVALSTR 509
Oy 481 DVTDLKVPKPVLDQYCKEFTAPVTYVLDKDKELKEOKLEAKYVRELVLNAGGEAEN 540
Db 510 DVTDLKVPKPVLDQYCKEFTAPVTYVLDKDKELKEOKLEAKYVRELVLNAGGEAEN 569
Oy 541 YVIVVLTKAGKEKATTLALTELKAPGAFSKFEYRGIDTELDRKYVTEENOKNATYVSLPV 600
Db 570 YVIVVLTKAGKEKATTLALTELKAPGAFSKFEYRGIDTELDRKYVTEENOKNATYVSLPV 629
Oy 601 DANGVLKCAEAAELKYTTTNEKGKVDATDAQVYVONNSVITVGGCAAGETKYTVVL 660
Db 630 DANGVLKCAEAAELKYTTTNEKGKVDATDAQVYVONNSVITVGGCAAGETKYTVVL 689
Oy 661 DDKLTITTSFKVVDTPAKGLAVEFTSTLKEVAPNADLKAALLNILSDGVPAATTAKA 720
Db 690 DDKLTITTSFKVVDTPAKGLAVEFTSTLKEVAPNADLKAALLNILSDGVPAATTAKA 749
Oy 721 TASNVEFSADTNVAENGTGAKGATSIYVKNLTVVKGDEQKVEFDKAVOYAVSIKEA 780
Db 750 TVSNVEFSADTNVAENGTGAKGATSIYVKNLTVVKGDEQKVEFDKAVOYAVSIKEA 809
Oy 781 KPATK 785
Db 810 KPATK 814

```

RESULT 2

S-layer protein precursor - Bacillus licheniformis
 C:Species: Bacillus licheniformis
 C>Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 15-Oct-1999
 C:Accession: J04930
 R:Zhu, X.; McVeigh, R.R.; Malachuk, P.; Ghosh, B.K.
 Gene 173, 189-194, 1996
 A:Title: The complete nucleotide sequence of the Bacillus licheniformis NM105 S-layer-en
 A:Reference number: J04930; M01D:97082965
 A:Accession: J04930
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-874 <RHU>
 A:Cross-references: GB:U38842; NID:g1055336; PIDN:AAC44405.1; PID:g1055337
 A:Experimental source: strain NM105
 C:Comment: This protein is a glycoprotein. It functions as protective coats, molecular s
 C:Genetics:
 A:Gene: alpA
 C:Superfamily: S-layer repeat homology
 F:1-29/Domain: signal sequence #status predicted <Sig>
 F:30-874/Product: S-layer protein #status predicted <Mat>
 F:157-209/Domain: S-layer repeat homology <SLR3>

Query Match

19.6%; Score 760; DB 2; Length 874;
 Best local similarity 31.9%; Pred. No. 8, 6e-25;

Matches 268; Conservative 114; Mismatches 327; Indels 130; Gaps 34;

```

Oy 1 AGTTPDPVADHMGIDISINLVKGAIVKNDKGFEPGKELITRAEATMMAQIINLPIDK 60
Db 30 AGKSPDPVAGHMAEDSINLVKGAIVKNDKGFEPGKELITRAEATMMAQIINLPIDK 89
Oy 61 DAKPSPADSGOWYTPPIAAVEKAGVTKGTG-NGFEPNGKIDRSMSLVEAKKLDTKV 119
Db 90 NADPSFADANLWSSKVIYAAVEKAGVTKGTG-KNDFEPNGKIDRSMSLVEAKKLDTKV 149
Oy 120 NGTPATFKLETLNMKGKANIIVELGISVGTGDQWEPKKTIVKAAAPFIATDKQFG 179
Db 150 DGLTVTKFDLIRG-HMGEKANIIVNLGISVGTGKKEPKSVSRALAPFIATDKKY- 207

```

```

Oy 180 TEAAKVESAAVTTQKEVFKSK-----AVEKTKEDIKVTNKANNDRVLKVEVTLSED 233
Db 208 ---AKRPENDAVATVNAAPTEPQLTLTGKLNKLTAEDEVTL---EGNKALIALEA--SKD 258
Oy 234 KSNATVELYSNLAQOTTVVDNKKVKETEVAVAGSLKAKTI-----EMADQTVAAE 284
Db 259 GRSAAVTLSGKTIAPNKELEPVKVK--GNFTIVKVVYEVKKLRVQLTFEDDRAOAAV--- 313
Oy 285 PTALOFTVDENGTE-----VVSPEGIEFY-----TPA-----AEKINAKGEITLAK 326
Db 314 -----FKLDEKGNADIEVLIDIGHVYKFAVANNLDGTPANIFPGGTAESTGLANGIAE 368
Oy 327 GSTTIVKAVYKKDKQKVAESKEVKSAGAAVASISNMVVA-EQNKA-----DFTSKDF 379
Db 369 GKRYKVEQVTKRGITIVSNTGILEVKNLDAEATAIKDVFPAVDTRAGVYAKPLSGTDF 428
Oy 380 KONK--VYEGDNA---YVOVELKDQFNATYTGKVEVESLNTTEVAVYDKATGKTVLSAGK 435
Db 429 TLNKSITLVAGEKAGIHRVVAQINKENKVVDPASISLKSNNP--GVLSVKNGETKAEPAAGS 486
Oy 436 APYKVTYVD-SKGAALVSHVEIEAFPAOKAMKDILEKTVALSTRDVTDLKVPKPVLDQ 494
Db 487 ATLTVKGVDTKTFDFVVKTDTRKLTTVKANPD-----QLKVYDGKELPVTFTTQD 538
Oy 495 YGKEFTA-----PVT--VKVD-----KDKG-----ELKEOKLEAKYVRELVLNA 533
Db 539 YGDPFGANSGAIKEVEPQTGVVVLVDTTNEBSIGTSSIKVKGENVAGCTIHFQNP-NA 597
Oy 534 AGQEGAGNYTVVLTAKSGEKEKATTLALTELKAPGAFSKFEYRG--LDTELDRKYVTEENOKN 591
Db 598 SGEIGSLHAYEVTKSNIGHAP---RIEL-----YSKAGKGGAADTTLGAGNTVAYQLS 649
Oy 592 AMTVSVLPVDANGVLKGAEAAELKYTTTNEKKEVEDATDAQVYVONNSVITVGGCAKAG 651
Db 650 NYTTEGVYADAADLAGYEFRVGMNDKIASAKIEGKTLKVTGKTAGV--TDVILTRKDGATAG 707
Oy 652 ETYKVVYLDGKLTTHSFVVDTPAKGLAVEFTST-----LKEVAPNND--LKAA 703
Db 708 HA--TIYTOENIDITTSKFEVDVE-----QFERKNYIDRVLDVVSDDQVNLNGI 758
Oy 704 LNILSDGVPAATTAKATASNVEFSADTNVAENGTGAKGATSIYVKNLTVVKGDEKE 762
Db 759 KLNI-STENKVRIVDESTBOGKYVLDNRNDNATDGNV-ALGYVTAVKSNIDTVYSKEGND 815

```

RESULT 3

S-layer protein precursor - Bacillus stearothermophilus
 C:Species: Bacillus stearothermophilus
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14850
 R:Jarosch, M.; Egelseer, E.M.; Mattanovich, D.; Sleytr, U.B.; Sara, M.
 submitted to the EMBL Data Library, April 1999
 A:Description: Nucleotide sequence of the coding region of sbSc, the S-layer gene fro
 A:Reference number: Z18239
 A:Accession: T14850
 A:Status: preliminary; translated from GB/EMBL/DBD1
 A:Molecule type: DNA
 A:Residues: 1-1099 <JAR>
 A:Cross-references: EMBL:AF055578; NID:g4581955; PID:g3025826; PIDN:AAC12157.1
 A:Experimental source: ATCC 12980
 C:Genetics:
 A:Gene: sbSc

Query Match

11.5%; Score 445; DB 2; Length 1099;

Best local similarity 24.6%; Pred. No. 1, 8e-11;
 Matches 223; Conservative 136; Mismatches 331; Indels 218; Gaps 41;

```

Oy 44 AEAATMMAQIIN-----LPIDKDAKPSFADSGOWYTPPIAAV 81
Db 28 SQAATDVATVVSQAQAKQAKMKEAVYTTTSHVTFTGQFPDIKIDVYAAVYAKK-QAYANAAVAV 86

```

[illegible]

Query Match	10.4%	Score 402.5	DB 2	Length 1176
Best Local Similarity	23.1%	Pred. No. 1.2e-09		
Matches 212	Conservative 139	Mismatches 374	Indels 191	Gaps 36
<p> A:Molecule type: DNA A:Residues: 1-1176 <BOW> A:Cross-references: GB:M8361; NID:9341911; PIDN:AAA50256.1; PID:9556012 C:Superfamily: S-layer repeat homology F:93-147/Domain: S-layer repeat homology <SLR2> </p>				
QY	16	DSINLVKRGAVKGDKGMFEFGKFLTLTAEAATMAAQLNLPIDKDAPSPFSDSGQ-WY	74	
Db	45	EAVQSLVDVAGVYQGGANGNFNPLKTRISRAEAATITPNALELEAGDV--NRFDVAKADAWY	102	
QY	75	TPFIAAVEAGATGK--TGNGFEPPNCKIDRVSMASLIVAEYKDTGVNCTPATPKFDLETL	133	
Db	103	YDAIATVYENGFEFEVSATFEPAPNQLTRSEAAKILLVDAFELEGED--LSEFADASTV	159	
QY	134	N-WGKEKANIIVLELISVGTG-----	157	
Db	160	KWMASTYLEIAANANGVINGSEANGKTNLPNAPITRQDPFAVVFSTIENVDATPKVDIE	219	
QY	158	-----PKTVYTKAAAOFLAKTDKQFG-----TEAAKVESAK	189	
Db	220	VVDAKTLNVTLSDDGKRETVTLEKALEPPKKEFEVTPKIKIDVEYKAKVFVVTATAVKSVS	279	
QY	190	ATTQKQVEKKFKAAXELTKEDIKYTNKANNDKVLYKEVLTSEDKRSATVELYSLLAKQ	249	
Db	280	ANLLEEVAFEPFGVYDKETAEAD--AANYALSKGKTISVLAADKRTATVTLTDLKLNKK	337	
QY	250	TYTVAVN--KVQKTVAAVAGSL-----FAKTIEMADQTVADPEP---TALQF	290	
Db	338	AAIAGISNKKAGDKELINKNVEFTAVNDKIPREVTEVKSIGTAAKAVTILSEPPENLSSTNF	397	
QY	291	IVKDE-----NGTEVVSPEGIEFVFPAAEKINAKGEITLAKGSTIVKAVYKD	339	
Db	398	TLDGKAFEGNVMVAGNKTIVILTPYSSSLSGVGDKCLTVSGAKDPAGEVSLSTHEPK--	455	
QY	340	GVVAESKEKVKASGAAVASISNMTVAEONKAD-----FTSKDKPKNKKYEE--GD	389	
Db	456	---VEDEKAPFTVATATALETTLTFLFSDDIDMDIVKASNYWYKSGDSKESSEFERIAD	512	
QY	390	NAYVOVELKDPQNAATYTGVEY-----ESLTEVAAVNDKATGKTVLSAGCAPKVTVK	443	
Db	513	NKTKRY--FKGSKITLPTGTGVVDYVEDIKDYSNKKIAKDTKYTVPEIDQTRFVAKKTAL	571	
QY	444	DSKG--KALVSHTVETIEAFQAKAMKDIKLEKTVALASTRDVTDLK--VKAPVLDOYKEFTA	501	
Db	572	DEKTIKVFPSKTVDESAIKTGTNYVKDDKDVSVDKTVDSKDSKSVIILYLSKVSYG	631	
QY	502	PATVAVLD--KQDKELKEQKLEAKYVNRELVLNAAQOEAGNTVYVLTAKSGKEAATLAL	560	
Db	632	ENTITIKNKYKDAFTKNNMTLD--TYGK--FTRSKQEGDEHVIYNADA--TAKRVVLKF	684	
QY	561	ELKAGGA---FSKEFEVRGLDT--ELDKVTEBENOKNMTVSVLP-----VDANG	604	
Db	685	DKKMAASLADYSNVLKINDTLQTLSEVATLVSNDATVYTTTFPAETIKDDVYVPASG	744	
QY	605	LVLKGA---EAAELKY---TTTNKGEKVDATDAQTVYONNSVITVQOGATAGETY--K	655	
Db	745	KAISSGKGVNVELQVMGVKQDTSGVHHKFKFNGSENKHTLSSVS--TPRLKLAIDKROYDAK	802	
QY	656	VVV--VLQGLKLTTHSEKVVYDTPATKGLAVETYSLSK-----EYAP	696	
Db	803	YAEALVDKRTVYVKESTYVNSA---AANAATSESHKIDSIOVNGSTVTVYKFEDEINT	857	
QY	697	NA---DLKALLNITISVDQVPAT--TAKRTAENFEVSADTVVVAENGVGAKGATSIYV	751	
Db	858	NASDDDLKAKNLKLVDIAGNESTNNTPLAIKGINLLDSVAPVVVVGEPVYDKETITFTFS	917	
QY	752	KNLVVKKQKEQKVEF	767	

Db 900 -VGEIKSDTGKATGAVSPVDDKFPSPFVSANGVYGASRKGFLLFDEDIKELNNS 958
Qy 755 -----TVVKDGR---EOKVEFDKAV---QVAVSIX 778
Db 959 AGIGATDLVYKNGSKTLEAGCIDYDAIDNKKITVTLK 995
RESULT 7
A56143
surface-array protein homolog sapA2 - Campylobacter fetus (strain 82-40 LP3)
C:Species: Campylobacter fetus
C:Date: 03-Nov-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: A56143
R:Dworkin, J.; Tummuru, M.K.R.; Blaser, M.J.
J. Bacteriol. 177, 1734-1741, 1995
A:Title: A lipopolysaccharide-binding domain of the Campylobacter fetus S-layer protein
A:Reference number: A56143; MUID:95204338
A:Accession: A56143
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1109 <DMO>
A:Cross-references: GB:S76860; NID:913763; PIDN:AA83871.1; PID:913764
A:Note: this gene appeared to be silent in strain 82-40 LP3
C:Genetics:
A:Gene: sapA2

Query Match 6.7%; Score 259; DB 2; Length 1109;
Best Local Similarity 21.2%; Pred. No. 0.0012;
Matches 202; Conservative 122; Mismatches 316; Indels 312; Gaps 46;
Qy 11 DHMGDSINYLVEKGAAGNDKGFPEPKELTRAEATMAAQLINLPIDK----- 60
Db 191 DEAGLTKIALTTENDITTEGEGDLISGVYGTAAEST-----LN-PGDKIDGAGNDVL 243
Qy 61 --DAKSPADSGQWTPPTFAAVERKAGVIG--TGNGFEPNGKIDRVSNASLVE----- 111
Db 244 KYDLKNNFGLKDDGY---IKNIEKLSLTVSSVSNRIFDAKG-IDGLQTVALSGEKISV 299
Qy 112 ---AYRLDTRKVNCTPATKFK----- 132
Db 300 TTNLANIVDVEVNGFKGTNPNVDSITADKVLDSADYONLKVNGVGAKGASVALTADKIET 359
Qy 133 LNN-----GKERANILVE--LGISVGTGDQWEPKKTVTFAEAQFIKTDKOGTE 181
Db 360 LNLNTGSGSFVSADVASISVKGANLISLATG-----AKTTT-LDASSFGALDADLSTS 413
Qy 182 AAKVESAKAVTQKVEVKSFAVEKLEKEDIKYT-----KNAANDKVLV----- 225
Db 414 A-----SVTS---IKGGNGNDKITIKDVAVNAIDGAGNDELVIKGSTADTLOPTL 462
Qy 226 -----KEVTLSEDKRSATVEL-VSNLAAK-----QTVTVVNVKVGKTEVA 264
Db 463 TNEKYTIDGNTKDLTSLKKAQSVTELSFKNIKATVYTESNGVNEYNIILANNATDKAVT 522
Qy 265 VGSLEAKTIEMAD-----QTVVADEPTALO-----TVKDENGTEV- 300
Db 523 INDESLKTIHFSDVDKGSVAAKGIVADKATELTINSKVTLASDAVVAQOANATKIDI 582
Qy 301 -----VSPGIEFVTTPAAEKINAKGETITLAKGSTYTKAVYKKKGKVAASKEKVS 352
Db 583 NNAKDTGVLTLGGVAKLTDLT--VNNKGAPALTGANATDLDV-----KNLSVN 629
Qy 353 AEGA-AVASIS-----NMTVAEQNKADF-----TSKDFQONNKVYEG 388
Db 630 TBEAFIATATSLKLNINLSNGVSADLNSVNTGATLALLEANINVSGEFKICTTAKG 689
Qy 389 DNAYVOVELK-DQFNAVTTGKVEYESLNTFAVAVDKATGVTY-----LSAGKA 436
Db 690 D-----VDFIEHVGALTLCATITSSGTNASV-IISATGNVTLGAVSAGQNLTLNAGNT 743
Qy 437 PVKVTYKDKSGKRL-----VSHTVLEAFQAQAMKDIKLEKTNVALSTADYVDLKKAP 490

Db 744 LGNITIGALGDIIVSDLVGGLCTINSDA-----NNKVSITSNEVT----- 784
Qy 491 VLDQYKEFPAPYTVKVLDDGKELEKQKLEAKYVNEVLVNAAGPAGVTVLITAKSG 550
Db 785 ---VGSSEISKNNVEITAAAGGTDILNAQVIGGAAADALITIGKD-----DQITTA-SG 835
Qy 551 EKAKATLALELKAPGAFSKFEVRGID-----TELDKYVTEBNKNAVTSLVPDA- 602
Db 836 DLSG-GFLTLTLEATKLSLSDISGVKGTGNVAIELKAV-----QGNKTDVSVQSGDAA 890
Qy 603 -----NGLVLKGAEEAELVYTTNKEGKEVDATDAQV-----V 636
Db 891 EQITYSAAISLTDIKISGDIGAGANTITVPDPAADLKTIDISLSATGCTLASTITLV 950
Qy 637 QNNSVITVGGAKAGETKYTVVLVDGKLITTHSPKVVDPPTAKGLVFEFTSTSKREVAP 696
Db 951 NANTATISVKGSLGADT--ITVVSANKAVAILDGK-DTDAID-----KVDYSRTISPSKN 1002
Qy 697 NADLKAALL--NILSYDV--PATTAKATAVNEFVSADTVVAENGTVG 742
Db 1003 DASIKADLVISITNALSGDQIVLKGATSIKDRGD-----LSEGANILALGKILG 1050

RESULT 8
A34355
cell surface protein precursor - Acetogenium kivui
C:Species: Acetogenium kivui
C:Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 08-Oct-1999
C:Accession: A34355
R:Petters, J.; Peters, M.; Lottspeich, F.; Baumeister, W.
J. Bacteriol. 171, 6307-6315, 1989
A:Title: S-layer protein gene of Acetogenium kivui: cloning and expression in Escheri
A:Reference number: A34355; MUID:90036724
A:Accession: A34355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-762 <PEPT>
A:Cross-references: GB:M31069; NID:q141844; PIDN:AAA21930.1; PID:q141845
C:Superfamily: S-layer repeat homology
F:32-85/Domain: S-layer repeat homology <SLR1>
F:96-150/Domain: S-layer repeat homology <SLR2>

Query Match 6.6%; Score 255.5; DB 2; Length 762;
Best Local Similarity 22.7%; Pred. No. 0.00099;
Matches 177; Conservative 89; Mismatches 275; Indels 239; Gaps 35;
Qy 5 FPDVPADHMGIDSINYLVEKGAAGNDKGFPEPKELTRAEATMAAQLINLPIDKRAK 64
Db 96 FKDPVPOHMAVVGQINILAYKLGIAQGVNGKFPDPSLETRYQALAFLVRLGFI--KDL-- 150
Qy 65 SFDPSQGWTPPTFAAVERKAGVIGTGNGFEPNGKIDRVSNASLVAAYKLDKFKVNGTPA 124
Db 151 -----DMFYGIYAKQDDGLVHGLNLA--NGLYIRGLALILDBALEV-----PM 194
Qy 125 TKFKDLETLNMGKEKANIIVELGISVGTGDQWEPKKTVTFAEAQFIKTDKOGTEBAK 184
Db 195 VKYVD-----GKE--VLGEPLIS-----KVATKAERY-VIATNAQDSVEEGK 234
Qy 185 V-----ESAKAVTQKVEVKSFAVEKLEKEDIKYTNKNAANDKVLVKE-----VTLSDKR 235
Db 235 VAVLDKDGKLTITINAGLVDESEYLGK---KVITYSSEFDPVYVAVAGNDVVSFTEGQD 290
Qy 236 SATVELYSNLAQKOTYVVDN-----KYGK-----TEYAV-----GSLGA 270
Db 291 SVGTYYKNDKNKTAIVDDNAYVLVNGYLTAKVSKVYVEGAEVYTIINNNTLYVNGSYDN 350
Qy 271 KTIEMADQTVVADEPTALOFTVADENGTEVVSPEGIEFVTTPAAEKINAKGETITLAKGTS- 329
Db 351 STI-----YNDVQSGDKYLRNDSN-----YELKGTVTVYTGAVSK 385
Qy 330 -TVVKA-----VYKKDKGV-----AESKEVYVSAEGAASISMWTV 366

```

Db      386 VTDIRKANDYIYGYQYDVNNGNVGTIVYVVRNOVGTGTEKSVSGSTYKASIDNV--YTV 444
      367 AEONKAD-----FTSKD-----FKONKRYEGDNAYVOEL 397
      445 ADNNAVNNQLEPGKRVYILNKDNYVIGISSTTTTANVNAIFREKSDPFTAMFAKVLIL 504
      398 KDO---FNAVTTGKVEYESLN---TEVAVDKATGRVYL-----SAGKAPVK 439
      505 PDAAEKVFADAVSD--YYDRVNLAEGTIVYTVDANGKLANDIORANDOPSSASAYKADAK 562
      440 VYVXDSGKALVSHVTEIE--AFQAKMKDKIKLEKTNVALSTKDVTDLKVAPLDQY-- 495
      563 VLREGSTTYTIDMTVTLNNTSDGFKALKLTDL-----KATNLNWKI--VADNNV 612
      496 -----GKEFTAPVYKVLIDKDKGKLEOKLEAKYVNRVLVNAAGOEAGNYTVLTA 548
      613 AKYVVFNNASVSTTSTTVY-----AYVTGADVYVNGSTFTRLTVLENGQ 658
      549 SGEKEAKATLALBKAPGAFSFEVRGLDTLDDKYVTEENKNAMTVSVLEVDANGVLK 608
      659 TKTYDANAOQLATNY-----THKAVVLTLTNKKIANI--ALPTVASGVKLT 701
      609 GAFAELKVTTNKEGKQVND---ATDAQVTYQNNSVITVGOGAK---AGETYKTVV 659
      702 NIDQANLRTTDTNKKYLLDPTNFIVDTNGNINLGLSDITKDTGVVLYNDGVKVFVIEIV 761

RESULT  9
140468
C:Species: Bacillus stearothermophilus
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Oct-1999
C:Accession: 140468; S34365
R:Kuen, B.; Sleytr, U.B.; Lubitz, W.
Gene 145, 115-120, 1994
A:Title: Sequence analysis of the sbas gene encoding the 130-kDa surface-layer protein
A:Reference number: 140468; MUID:94320770
A:Accession: 140468
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1228 <RES>
A:Cross-references: EMBL:X71092; NID:g312729; PIDN:CAAS0409.1; PID:g312730
C:Genetics:
A:Gene: sbas

Query Match      6.4%; Score 248.5; DB 2; Length 1228;
Best Local Similarity 22.4%; Pred. No. 0.0036;
Matches 201; Conservative 109; Mismatches 309; Indels 279; Gaps 44;

      55 NLPIDKDAKPSFADSGQWYTPFIAAVEKAGYIKGTGNGFEPGKIDRYSMASLVEAYK 114
      323 DVSLSLDGKTTIVDAS---TPENNTEYKVVYKGIK---DANGKEFKEDATFFKLRNDA 375
      115 LDTKVGCTPAT-----KPKLETIN-----WGK---- 137
      376 VVTGVGCTVNTNNTSVNLAAGTFDODTLVPEDKLAPETVSSVNTITDVEYGRIPY 435
      138 ----EKANILVELGISVGDOME---FKTYT--KAAEAOFIATDQOFGTEAKVES 187
      436 IASTSGSTITILKEALVYGKQYKLAINNVTLLGYNAEYELV-----FTNANASPTV 489
      188 AKAVTT-----QKEVYKFSKAVEKLTREDIKVT 215
      490 ATAPTTLLGCTTLSTGSLTTVMWCKLAGVNEAGTYVPGQFTTTPATKLDESLADNFVL 549
      216 NKANNDKVLVKEVTLSEDKRSATVELYSNLAAGOTTVVNNKAGKE--VAASGLEKTI 273
      550 VERESGTVVASELKYNADAKAMTVLVPRADLKENTYIOIKIKGKSGKIELGTVNEKTY 609
      274 EMADQVVADEPRLAQFTVVD--ENGTEVVSPEGIEFVTPAAEKINAGETTLAKGISTV-- 330

```

```

Db      610 EFKTQDLTA--PTVIVYSKNGDAGLKVT--EAOEFTVKESENINENTATTVSGSTITYG 665
      331 -----TVKAYYKKDG-----KVASESK----- 348
      666 QVAVYKAGANISALTASDIIPASVEAYTGQDGTIKVYVANNQLENRNGYKLVYVGKATA 725
      349 -VKVSAGAAVASISMTVAEONKADFTSKDEKONKRYEGDNAYVOELKDOGNATYTG 407
      726 PVKAAANANLTATNYIYFTTEGQ-DVYAPTV---TVFKGDS-----LKDA-DAVTT- 773
      408 KVEYESLNTVAVVDKATGKTVLSACKAPYKTVY-----KDSKGRALVSHVTEIEFA 461
      774 -----LTNVDAGQ--KFTIQSEELKTSIG--SLVSGKTVTEKLT 809
      462 OKAMKDIKLEKTNVALSTKDVTDLKVKAPVLDQYKKEFTAPVYKVLIDKDKGKLEOKLE 521
      810 NNGWVD---AGTGTTVSV-----APKTDANGKVAVAVTLGLDNNDKDAK-LRLV 856
      522 AKYVNRVLVNAAGQ-----EAGNTYVLTAKSGEKEKATLALBKAPGAFSK 570
      857 VDKSSTDGIDAVGAVNIKEKDILIRYNSWRHTVSVKKAADKQDQ-----NSAAPP- 908
      571 FEVRGLDTELD---KYVTEENKNAMTVSVLPVDANGVLKGAFAELKVTTNKEGKQV 627
      909 -----TSTAIDTYSLVEFNE-----TDLAEVKRENIVVDAAGNAVAGTVTALDGS-- 956
      628 DATDAQVTYQNNSVITVGOGAKAGETYKTVVLDG-----KLITTHSFKVYVDPAP 677
      957 -----TNKEFVPTPQELKAGTVYSVTI--DGVVRDKVGNITISKYIT--SEKTVSAMP 1003
      678 TAKGLAVEFTSISLEVPANMDLKAALLITLSDVGP-ATTAKATASVNEEVSAD-TNVV 735
      1004 T-----LSSISIDAGANVD-RSKTITIERSDSVPNPNTTLKKRADGSIFFNTLYVNN 1055
      736 AENGT---VGAGAT--SIYVKNLTVYKDGK-----EOKVEPDKAVQVAVSIEKAPKA 783
      1056 NENKRYKIYFHNGVILDEFTQYELAVSKQFQGTGTDIDSKVTF---ITGSVATDEVKPA 1110

RESULT 10
S33441
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abse
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SM>
A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAAS0714.1; PID:g298032

Query Match      6.3%; Score 245.5; DB 2; Length 1822;
Best Local Similarity 21.2%; Pred. No. 0.0082;
Matches 201; Conservative 120; Mismatches 345; Indels 281; Gaps 41;

      15 IDSINYLVEGAVKGNDKGNFEPGKELTRAATAATMAAQIINL-----PIDKD 61
      914 LDALNKL--EKDSE--TKAIDANPNLTPPEKAKALAKVLELVNNAESDILSKPRPTVQ 970
      62 AKPSFADSGQWYTPFIAAVEKAGYIKGTGNGFEPGKIDRYSMASLVEAYK 108
      971 AVEDEKADKD-----LAKVEIQAAADGAKKGIENPNLTPPEKDAKKAVEDAVKAYATD 1023
      109 LVEAYKLDTKVNGCTPATPKFQDLETLNW---GKEKANILVELGISVG-----TGDQW 136
      1024 AIDKASPTVEVDRTATSGVAAIDAEEFKATQOKAKNNIAEABESAKKAIDNPMLTPDEK 1083
      157 EPKRYTVKAAEAOFIATKDTQOFGTEAAKVASAKAVT-----TQKVEK----- 199

```

```

Db 1084 ESKANVAEAAKVATPAALIDKSTPDVAOVEDDKGPAALINLTTRAKADAKGVIAAKLADIEIK 1143
QY 200 -----FSAVE--KLTIEDIKYTKANKANDKVLKVEVTLSEDKRSATVELY----- 242
Db 1144 KLEDKQAEAEKAIDAOSTMTNEEKALAKKALODVVDKGAELEDAARVATNIEHATYTEK 1203
QY 243 ---SMLAKQRYTVDVNVKVGTEVAVGSLEKATJEMAOQTVADPEPTLOETVK----- 293
Db 1204 AKAAELAGEKSILT-DTGKEARDAVEL---AKDKELAKEARIRTEEBEATKIVKEKLAEDTR 1258
QY 294 -----DENG-----TEVWSP-----EGIEFVTPAEAKINAKGEITLAKGST 330
Db 1259 KAIEDNPMLSEDDKQAEIKKLTDAVAKTLATIRBNADKRTOGEAEKAQGLADLEKAKETQK 1318
QY 331 TVK-----AVYKKDKGVVAESKEVK-----VSAECAAVASISNMTVAOONKADFT-- 375
Db 1319 IADKKAIDRLTILYKDGELKATKODAKNKRIAKDAAAEEAIAISNPLDAE--KRTFEDA 1376
QY 376 --SKDFKONN-----KYEGDNAYVOVELKDOFNAV---TTGKEYEESLMTVEAV- 420
Db 1377 VDAEVAKANDAIISAATSPADYOKKEBDAGVAAIAEVLDAQAODAKNKIAKAAAKAEKAI 1436
QY 421 -----VDKATGKYT-VLSAGKAPYKVTVKDSKGALVSHTEVEIAFAQK 463
Db 1437 SNPMLTDAEKRTFTDVADEVAEKANDAIISAATSPADYOKKEBDAGVAAIAEDV-LDAAKOD 1495
QY 464 AMKDIKLEKTVNLSTKRYTDLKAKAPVLDYOKKEFT-----APYVYKVLDDGKELK 516
Db 1496 AKNKIAKE-SDAASAIIDN-----PMLTDAEKESAKKAVADADAKAATDAIDASTSPE 1548
QY 517 EOKLEAKVYN--RELVLNAAOGAGNYTVVLTAKSGEKEAKATLLELAKAGAFSKEFVR 574
Db 1549 AQSADKCKVGSIAQVLDLAAQODAKN-----KIAKEVAAKAEKIDANP-----NLS 1594
QY 575 GLDPELDKYVEENOK---NANTVSVLEPYDANGVLKKA-----EAELEKYTTNKE 623
Db 1595 DAEKESAKKAVDADAKATTDIDAISTSPVEAOSADEKCKVGSIRQVLDLAA--KODAKNKI 1652
QY 624 GREVDATAQOVTYVQNSVITYGOGAKAGETKYVTVLBDGKLITTHSFVVDPTAPAKGLA 683
Db 1653 AKESPAAKSAIDANPLTDAEKESAK-----KAVD----- 1682
QY 684 VEFSTSLKEVAPNADLKAALNLTISVDGVPATTAKATASNYE-----FVSAD----- 731
Db 1683 -----ADAKAA-----TDAIDASTSPVEAOSADEKGVCAIKDILDAKQ 1722
QY 732 ---TNVAENGTVGAKGATSIYVKULTVYKDGKE-OKVEFDKAVQYAV 775
Db 1723 DAKNKIAE---AESAKSVIDSNPMLTDAKAEKASEIDKAVAEVAI 1765

```

Query Match	6.2%	Score 241;	DB 2;	Length 1128;
Best Local Similarity	25.5%	Pred. No. 0.0067;		
Matches 191; Conservative	90;	Mismatches 353;	Indels 116;	Gaps 34;

RESULT 11
T30296
R27-2 protein - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C:Accession: T30296
R:Otsu, K.; Donelson, J.E.; Kirchhoff, L.V.
Mol. Biochem. Parasitol. 57, 317-330, 1993
A:Title: Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amino
A:Reference number: Z20813; M0ID:93165082
A:Accession: T30296
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1128 <OTS>
A:Cross-references: EMBL:L04603; NID:g365171; PID:g1256742; PIDN:AAA6494.1
A:Superfamily: neurofilament triplet H protein

QY	99	KIIIVSNASLIVE----	AYKLDTRKVNTPPTKRDLETLMMGXEKNIIIVELGISTGSD	154			
Db	15	KLDINRRTKLEEBERNQATKAHKAEEKKTRÖRLITFE--	SENINLKRPDVAISND	72			
QY	155	ÖMEKRTV-----	TKRAQOFIATDKQFSTEARV-----	BSAKVTTÖKVE 197			
Db	73	KKKSEIAPKTDEVEKÖRAAEAKAV-ETEKÖRAAEATKVAEAKRKAAEAKVETEKÖR		131			
QY	198	VKESKAVEKLTEDIKYTNKA--NNDKVLKVEVY--LSEDKRSATVELYSNLAKÖTYITVY		255			
Db	132	AAEATKVAEEKÖKAAEAKAAVETEKORAAEATKVAEEKORAAEEMKVAEEKORAAEA		191			
QY	256	NKVAKTE-----	VAVSLUEAKTTEMAQOTVAADEPITALQTYVDEKENGTEVSPGIEPPT	310			
Db	192	TKVAEEKÖKAAEATKVAEEKÖKAAEATVVAEEKÖ----	KAAEATKVAEE-----	KO 242			
QY	311	PAAEKIN-ANGELITLAKGTSTTVKAVKK--	DEKVAAESEKVEKVSAGAAVMSISMTVA	367			
Db	243	KAABATVVAEEKÖKAAEATKVAEEKÖKAAEATKVAEAEKÖR--	AAEATKVAEAEKÖKAA	301			
QY	368	EÖNK--ADETSDEKÖNNKYYEGD-----	NAYVOVELKQFNATVTKGVEYESIN----	TE 417			
Db	302	EATKVAEEKÖKAAEATKVAEEKÖKAAEATKVAEEKÖKAAEATKVAEEKÖKAAEATK		361			
QY	418	VAVVDKATG----	KVYVLSAGKAPKVTYVDSKGKALVSTVTELEAPQAMMDIL----	470			
Db	362	VAAEAEKÖKAAEATVVAEEKÖKAAEATKVAEAEKÖKAAEATKVAEEKÖKAAEATVVAEA		421			
QY	471	EKTVALSTKVDYDLKAKAPALDQYG--	KEPTAPVYTKVLBDQGELEKÖKLEAKVNR	527			
Db	422	EKÖKAAEATKVAEEKÖKAAEATVVAEAEKÖKAAEAT--	KVAEE-----	KÖKAAEA 471			
QY	528	ELVINAQÖEGNTVVLVFAKSGEKAATLALKAPGAFSFEYRGD----	TELDKY	583			
Db	472	TKVAEEAEKÖKAAEATKVAEAEK--	KAAEATKVAEAEKÖKAAEATKVAEAEKÖKAAEATRV	530			
QY	584	VTEENÖKNAIMVSVLPDANGVLVKGAAEALVYTTNNKGEKVEDVTDQVYVQNNSVIT		643			
Db	531	AAEAEKÖKAAEATKVAEEKÖ----	KAAEAT--	KVAEEKÖ--	KAAEATK--	-----	VA 573
QY	644	VGOGAKAGETVKVYVVDGLITTHSPKVVDTAPTAKGLAVEFTSTSLKEVAPNADLKAA		703			
Db	574	EAEKÖKAAEATKVAEAEKÖK-----	AAEATKVAEAEKÖKAAEAT----	KVAEAEKÖKAA	623		
QY	704	-LNLISVDGVPATTAKATASNVEEVSADTNVAENGTGAKGATSI-----	YKNLVY	756			
Db	624	EATKVAEAEKÖKAAEATKVAEAEKÖKAAEATKVAEAEKÖKAAEATVVAEAEKÖKAAEATK		683			
QY	757	VKDGEKÖKVEEDKAVÖYAVASIKI-E	AKATK	785			
Db	684	VAAEAEKÖKAA--	EATKVAEAEKÖKAAEATK 711				

RESULT 12
T34418
hypothetical protein F12F3.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #extL_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohlmann, P.
submitted to the EMBL data library, July 1998
A:Description: The sequence of *C. elegans* cosmid F12F3.
A:Reference number: Z21521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U08022; PTDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2, clone F12F3
C:Genetics:
A:Gene: CESP:F12F3.3
A:Map position: 5


```

RESULT 14
B33856
hypothetical 80K protein - Bacillus sphaericus
C:Species: Bacillus sphaericus
C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 30-Sep-1993
C:Accession: B33856
R:Bowditch, R.D.; Baumann, P.; Yousten, A.A.
J. Bacteriol. 171, 4178-4188, 1989
A:Title: Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein
A:Reference number: A33856; MUID:09327128
A:Accession: B33856
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-745 <BOW>

```

```

Query Match 6.1%; Score 236; DB 2; Length 745;
Best Local Similarity 21.9%; Pred. No. 0.0063;
Matches 183; Conservative 98; Mismatches 260; Indels 296; Gaps 38;

OY 39 KELTPAEATATMAQILNIPIDDAKPFADSOGQVYTPFIAAVEKAGYIKTGNGFEE--- 95
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 78 KEAPPIKEATATLEVTYTLTFSEDDVMDIVKASNYW-----KSGSKKASFEFERIA 129
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 96 -----PNGKIDRVSMASLLVEAYK-----LDTKVNGTPATKFKEDLET 132
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 130 DNKYKFVEFKGAEKTLPTGKVD-----VYEDVKDYSDMKIAKDTKRYATPEI----- 176
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 133 LNMGKEKANIILVELDISVGTGDQMEPKTYTKAEAAQFIAKTDKPGFGEAAVEKAKVT 192
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 177 -----DQTRP-----EVRKVT--VD 190
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 193 TQKVEKFSKAVEKLTLEKEDIKVTNRKANNDK---VLVEEVTL-SEDKRSATVELYSNLAA 247
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 191 EKIIVTFESKTVDEKFTAE-KAGNYTIDKDGKVVSVKVIYDSDSKSVIIDLSKYS- 247
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 248 KOTYIVDVNKKGKTEVANAQGLEAKTIEIMADQTVVADDEPALOTFYKDKENGTELVVS-PGCI 306
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 248 -----VGENITITIKNVK-----DATKLNMTIMDYGKFTFSDKEGP 283
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 307 EEVTPAEKINKGETTAKGTSTYVKAVYKKDKGVKESKRVKVSAC-----AAV 358
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 284 KFEET---VINAD---AKAKVYLVKFNKKKDAASLADSSNYIVRIGDTLOTLDVDAIT 334
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 359 ASISNMVTAQONKADFT-----SKDFKONNKVYEGDNAVYQVELKDFQNAVYTGKVEYE 412
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 335 LSVSN-----DATVYTIFFAETIKGNDVFA-----TGKTSKANVH 371
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 413 SL-----NTEVAVVDKATGKVTVLSAGKAPYKTVVYKSKGALVSHVEIFAEPQKAMK 466
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 372 ELQVILGVRDTSQNVHDKNGKDN-----IDLTVGTTK-----LAFMQ----- 409
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 467 DIKTEKTNVALSTKQVTD---LKVKAPV-----LDQKGEFTAPVYKV 507
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 410 ---IDKDDAKTTLAELVDKRYKYKVFSTYIKSASSNAFTSNHTKIDSIOVDGTSITYYKF 466
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 508 LDK---DGEKLEOKLEAKYVN-----RELVINAAQOEGANTYVVLTFKSGEKEKA 556
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 467 KDEITKDSDLVLVNLISKFVDVADNEGPRBQITSPPTNLLDVAAPVLDGEPVYKDAIT 526
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 557 TLAL---ELKAPGAFSKFEVKGIDTLELDKYVTEENOKNMTVSVLEVDANGVLKGAENAE 614
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 527 TTFSESLSKAGS-----DVL-----AND 546
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 615 LKVTYTNKGEKVDATDAQVYONNS---VITYGQAKAGETYKVTYVLDGLKITTHSEFV 672
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 547 LVTYTVS-DNMDLAISDTYVANNDKKQVITILSDKREATAKATVY-----KNAKTI 597
   || : || : || : || : || : || : || : || : || : || : || : || : ||
OY 673 VDTAPTAGLAVEFTSTSLKEVAPNADL---KAALLNLT---SYDGVPTAKATASNVE 726
   || : || : || : || : || : || : || : || : || : || : || : || : ||
Db 598 IDTSPPKNAIA-DESKTATADKVOTSDTIGENTAAALVKLNKAIDKRTKTLQVAVGTT 656
   || : || : || : || : || : || : || : || : || : || : || : || : ||

```

QY 727 FVSADTNVAENGTVGAKGATSIYVKNLTVVKDGKEQVEFDKAVQVAVSJKEAKPA 783

Db 657 KLDSTNFRAVNNAAAALVADLNTAK--TAVEGATYTLLEATDTSYTAALKAKAAVEA 710

<p> A:Accession: A48490 C:Species: Thermomanaeobacterium saccharolyticum C:Date: 03-May-1994 #sequence_revision 11-Apr-1997 #text_change 16-Jul-1999 C:Accession: A48490 R:Lee, Y.E.; Lowe, S.E.; Henricson, B.; Zeikus, J.G. J. Bacteriol. 175, 5890-5898, 1993 A:Title: Characterization of the active site and thermostability regions of endoxylan A:Reference number: A48490; MUID:93388520 A:Accession: A48490 A:Molecule type: DNA A:Residues: 1-1231 <LEF> A:Cross-references: GB:M97882; NID:g533366; PIDN:AAA21812.1; PID:g533367 A:Experimental source: Strain B6A-11 A:Note: mutation experiments support roles for Asp-537, Glu-600, and Asp-602; His-572 -495, Asn-559, and His-572 are supported by X-ray reference A55905 on another molecule C:Genetics: A:Gene: xynA C:Function: A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in xylans A:Pathway: xylan degradation C:Superfamily: Thermomanaeobacterium endo-1,4-beta-xylanase A; S-layer repeat homology A xylanase A cellulose-binding repeat homology C:Keywords: duplication; glycosidase; heat-stable protein; hydrolase; polysaccharide F:1-93/Domain: signal sequence #status predicted <SIG> F:34-1231/Product: endo-1,4-beta-xylanase A #status predicted <MAT> F:39-186/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA1> F:196-340/Domain: Thermotoga xylanase A amino-terminal repeat homology <TXA2> F:368-676/Domain: Streptomyces endo-1,4-beta-xylanase A homology <STX> F:683-854/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC1> F:858-1043/Domain: Thermotoga xylanase A cellulose-binding repeat homology <TXC2> F:1056-1109/Domain: S-layer repeat homology <SLR1> F:1115-1168/Domain: S-layer repeat homology <SLR2> F:1179-1231/Domain: S-layer repeat homology #status atypical <SLR3> F:495,572,600,602/Active site: Glu, Asp, Asn #status predicted F:537,539/Binding site: substrate (Asp, Asn) #status predicted </p>	
--	--

Query Match	6.0%;	Score 234.5;	DB 1;	Length 1231;
Best Local Similarity	34.3%;	Pred. No. 0.014;		
Matches 60;	Conservative	30;	Mismatches 78;	Indels 7;
				Gaps
QY	2	GKTEPDVADHWGIDSTINLYVERGAVKANDKGFEPGEKELTTRERATMAQILNLPJKD	61	
		: : : : : : : : : : : : : : : : : : : : :		
Db	1053	GVTFDDI-KNSMAKDALEVLASRIYEGMTDTQIEPNKVTITREFLAMLRLINK-BEQ	1110	
QY	62	AKPSFAD-SQGWTPPIAAVERAGVYKGTGNGFEPNGKIDRYSNASLVRVAKLDT--K	118	
	: : :	: : : : : : : : : : :		
Db	1111	YGSERDVSNSCDWANNLEAKYKAGLIEGSKARRNDSTITREPMALIMABAYEMLQYK	1170	
QY	119	VNGTPATPEKDLTL-NWGEKANILVELGISVG-TGDQMEPKVTYTKAEAKOFT	171	
	: : :	: : : : : : : : : : :		
Db	1171	ENIGATGSESDSKISDMKRVNVAANAQKILVNGEPPNNEAFKDIATREAAAIL	1225	

Search completed: March 23, 2002, 07:53:57
Job time: 3415 sec

Db 598 SGGGYSLSHVEYTKSNIGHEAP---RLEL-----VSKAGKGADPTTLCAGNTVAOLS 649

Qy 592 AMTVSVLPVDANGLVKGAEEALKYTTNKESKEVDATDAOVYVONNSVITVGOGAKAG 651

Db 650 NTTEGVVADADLAGVEFRVGNDKIASAKIEGKTLKVTGKTAGV--TDVILTRKDGATAG 707

Qy 652 EYTKVTVLDGKLTITHSFVVDTPAPLAKGLAVEFTSTS-----LKEVAPNAD--LKA 703

Db 708 HA-TTIVTEENIDITSKVEFDVEE-----QFENRKVNIDRVLDVYKSDKDDVINGI 758

Qy 704 LITLISVDGVPATATKATASNEFVSADTVVAENGTVGAKGATSIYVKRLTVYKDGKE 762

Db 759 KLNI-STEHKVRIVDESTEQGVYLDNRNATFDGNDV-ALGVTVAVKSDPTVSKEGND 815

RESULT 3

SLAP_BACSH STANDARD: PRT: 1176 AA.

AC P38537;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE SURFACE-LAYER 125 KDA PROTEIN PRECURSOR.

OS Bacillus sphaericus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

CC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxId=1421;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=2362;

RX MEDLINE=89327128; PubMed=2666389;

RA Bowditch R.D., Baumann P., Yousten A.A.;

RT "Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein from Bacillus sphaericus 2362 and of a related cryptic gene.";

RT J. Bacteriol. 171:4178-4188(1989).

RL - FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY OF PROTEIN WHICH COAT THE SURFACE OF BACTERIA.

CC - SUBCELLULAR LOCATION: CELL WALL.

CC - SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M28361; AAA50256.1; -.

DR PIR: A33856; A33856.

DR InterPro: IPR001119; SLH.

DR Pfam: PFO0395; SLH; 3.

DR PROSITE: PS01072; SLH DOMAIN; 2.

DR Signal: Cell wall; S-layer; Repeat.

FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 1176 SURFACE-LAYER 125 KDA PROTEIN.

FT DOMAIN 32 91 SLH 1.

FT DOMAIN 92 151 SLH 2.

FT DOMAIN 152 210 SLH 3.

SO SEQUENCE 1176 AA; 125225 MW; 1A9AA58EF433788C CRC64;

Query Match 10.4%; Score 402.5; DB 1; Length 1176;

Best Local Similarity 23.1%; Pred. 4.1e-09;

Matches 212; Conservative 139; Mismatches 374; Indels 191; Gaps 36;

Qy 16 DSINYLYEKAAGVGNCKGMEPEGKELTRAPALYMAOILMLPDKDAKPSFASDQGO-WY 74

Db 45 EAVQSILVADAVIIGDANGNENPLKTRASRAAATFTNALTELEAGDV--NFKDVKADAWY 102

Qy 75 TPPIAVERKAGVKG--TGNGEFPNGKIDRVSMASLVEAVKLDKRVGTATKFKDLETL 133

Db 103 YDAIAATVENGIFEGVSATEFAPRKQLTRSPAAILVDAPFELBEGD---LSEFADSTV 159

Qy 134 N-WGKERANILVELGISVGTG-----DQWE 157

Db 160 KPMKXSYLEILAVANGVJKSGANGKTNLPNAPITRQDFAVVFSRTIENVPATPKVKIE 219

Qy 158 -----PKRTVYKAEAOPIAKTDKQFG-----TEAAKVESAK 189

Db 220 VVDAKTLNVLSDGTEKETVLEKALPEPKETEFYKIKDVEYKAKVYVYVTTATAVKSVS 279

Qy 190 AVTTQKVEVFSKAVEKLTEDIEDIKVTKANNDDKLYVEVLTSEDKRSATVLSNLAQK 249

Db 280 ATNLEEVVEEDGTVDKETAD--AANYALKSGKTTISVSLAADNKTATVTLTKLNNK 337

Qy 250 TYTVDVN--RVGKTEVAVGSL-----EAKTIEMADQTVVADEF---TALQF 290

Db 338 ADAISISNVAKGDEINVKVNEEFAVDNKKIPEVTEVKSIGTKAVKVLSEVENLSSNFE 397

Qy 291 TVKDE-----NGTEVSPGIEFVTPAAEKINAKGEITLAKGSTTTVAAYVKD 339

Db 398 TLGKAYFGVVMGAGKRTYILPEYSSALSVDGDKLTVSGAKDPAGFVSIINSTHERK-- 455

Qy 340 GKVAESKEVYKSAEGAASISNMTVAEQKAD-----FTSKDFKONKRYE--GD 389

Db 456 ---VDEKREAPTYEATATLETYTLTFSEDIQMDTVKASNVYKWSGSKKASEFERIAD 512

Qy 390 NAVYVELKDOFNATVTGVEY-----ESLNTVAVVDKATGVTVLSAKAKVYTVK 443

Db 513 NKYKFV-FKQSEKTLPTGKQDVYVEDIKDYSDNKIADTKVTVPPEIDQTRPEVRKVTAL 571

Qy 444 DSKG-KALVSHITVEIEFAQKAMKDIKLEKTNVALSTKQVTLK-VKAPVLDQKRETA 501

Db 572 DEKTIKVTFSKTVGDESAIKGNTYVKDKDKVYSVKVTYVDSKSVIIDLYSKSVG 631

Qy 502 PVTVKVLVD-KDGKELKQKLEAKYVNRVLVNAAGQNGNTVTLTAKSGKEAKATLAL 560

Db 632 ENTITINVKDATLNTMTLD--YTGK---PTRDKEGPDVEHYINADA-KAKKVLKF 684

Qy 561 ELKAPGA---FSKFEVGRDIT--ELDKYTEENOKNMTVSLP-----VDANG 604

Db 685 DKRDAASLADYSNYLVKINDITQTLSEDAVTLVSNDAFTVTTFAETIGDDVVEFASG 744

Qy 605 LVYKGA---EAEELKV---TTTNKESKEVDATDAOVYVONNSVITVGOGKAGETV--K 655

Db 745 KAISGSKRVNNELOVNGVKDPTSGVHKFKFNGSENKTTLSSTS--TPLKLAKIDKDYDAK 802

Qy 656 VTV-VLQGLKLTITHSFVVDTPAPLAKGLAVEFTSTLK-----EYAP 696

Db 803 YTAELVDKRYKVKFSTVYNSA-----AANAFTSESKRIDSIOVNGSTVYVKRDEINT 857

Qy 697 NA---DKAALLNLSVDGVPAT--TAKATASNVEFVSADTVVAENGTVGAKGATSIYV 751

Db 858 NASDLDLKVLNLSLVDIAGNENSTNNPTLAIKAGINLDSVAPVYVGEVYVKEKITTFES 917

Qy 752 KNLTIVVDGKEQKVEF 767

Db 918 ENLTSVIGEVLSTDF 933

RESULT 4

SLAP_ACEKI STANDARD: PRT: 762 AA.

ID SLAP_ACEKI

AC P22258;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE CELL SURFACE PROTEIN PRECURSOR (S-LAYER PROTEIN).

OS Acetogenium kivui.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

CC Thermoaerobacter group; Thermoaerobacter.

OX NCBI_TaxId=2325;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X71092; CAA50409.1; -
DR EMBL: AX000218; CAB77069.1; -

DR PIR: S34365; S34365.

DR HSSP: P02766; 1ETB.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

DR Signal: Cell wall; S-layer.

Db 1004 T-----LSSISADGAVND-RSKTITTFSSVNPPTTLKKADSTSTNTLVNVN 1055
Qy 736 AENGT---VGAKGAT--SIYKNTLVYKDGK----EOKVEFDKAVOYVSIKAKPA 793
Db 1056 NENKTYKIVFHKGVTLDEPQYELAVSKMDPQTGIDISKVTF---ITGVSATDEVKPA 1110

RESULT 6

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

APU_THETU

RESULT 9		
CNA_STAU	CNA_STAU	STANDARD:
ID	CNA_STAU	PRT; 1183 AA.
AC	Q53634.	
DT	15-DEC-1998	(Rel. 37. Created)
DT	15-DEC-1998	(Rel. 37. Last sequence update)
DT	15-DEC-1998	(Rel. 37. Last annotation update)

DE	COLLAGEN ADHESIN PRECURSOR.	
GN	CNA.	
OS	Staphylococcus aureus.	
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacillus/Staphylococcus group; Staphylococcus.	
OX	NCBI_TaxID=1280;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-EMBL 574:	
RX	MEDLINE=92165839; PubMed=1311320.	
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,	
RA	Lindberg M., Hoeoek M.;	
RT	"Molecular characterization and expression of a gene encoding a	
RT	Staphylococcus aureus collagen adhesin.";	
RL	J. Biol. Chem. 267:4766-4772(1992).	
RN	[2]	
RP	ERRATUM.	
RA	Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,	
RA	Lindberg M., Hoeoek M.;	
RL	J. Biol. Chem. 269:11672-11672(1994).	
RN	[3]	
RP	COLLAGEN-BINDING DOMAIN.	
RC	STRAIN-EMBL 574:	
RX	MEDLINE=94032261; PubMed=8218209;	
RA	Patti J.M., Boles J.O., Hoeoek M.;	
RT	"Identification and biochemical characterization of the ligand	
RT	binding domain of the collagen adhesin from Staphylococcus aureus.";	
RL	Biochemistry 32:11428-11435(1993).	
RN	[4]	
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.	
RX	MEDLINE=974745225; PubMed=9334749;	
RA	Myers J.J., Patti J.M., Carlson M., House-Pompeo K., Teale M.,	
RA	Moore D., Jin L., Schneider A., Delucas L.J., Hoeoek M.,	
RA	Narayana S.V.L.;	
RT	"Structure of the collagen-binding domain from a Staphylococcus	
RT	aureus adhesin".	
RT	Nat. Struct. Biol. 4:833-838(1997).	
CC	-I- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO	
CC	COLLAGEN-CONTAINING SUBSTRATA.	
CC	-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL	
CC	-I- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS	
CC	IN THE REGION OF THE MEMBRANE ANCHOR.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isp.slb.ch/announce/	
CC	or send an email to license@isp.slb.ch).	
CC	-----	
DR	EMBL, M81736; AAA20874.1; -.	
DR	PDB; 1AMX; 24-JUN-98.	
DR	InterPro; IPR001899; Gram_pos_anchor.	
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE NEG.	
KW	Signal; Repeat; Transmembrane; Cell wall; 3D-structure.	
FT	SIGNAL 1 29	POTENTIAL.
FT	CHAIN 30 1183	COLLAGEN ADHESIN.
FT	DOMAIN 30 1157	EXTRACELLULAR (POTENTIAL).
FT	TRASMEM 1158 1177	MEMBRANE ANCHOR (POTENTIAL).
FT	DOMAIN 1178 1183	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 151 318	COLLAGEN-BINDING.
FT	DOMAIN 533 1093	3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT	DOMAIN 1093 1157	LYS-PRO-RICH (CELL WALL-SPANNING).
FT	DOMAIN 1151 1156	CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT	REPEAT 533 719	PROTEINS.
FT	REPEAT 720 906	B1.
FT	REPEAT 907 1093	B2.
FT	SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;	B3.

Query Match 5.5%; Score 213.5; DB 1; Length 1183;

Best Local Similarity 22.0%; Pred. No. 0.11;
Matches 199; Conservative 109; Mismatches 304; Indels 291; Gaps 47;

```

OY 19 NIVLEKGVKNDKMGEPGKELTRAEATMAQILNP-----IDKDA 62
DB 232 NYSQSAITDFEKAFF--PGSKITVDNTRKNTID--VTIPQGYGSYNSFKTKITNQ 287
OY 63 KPSFADSGOWYTPP-----IAAVERKAGVYKGTNGEPNGKIDRVSMASLL 109
DB 288 OKEFVNNQAOYIOEHGKEVEVNGKSPNHTVHNINANAGLEGTVKGELKVLKDDKTRKAPIA 347
OY 110 VEAVKLDTRVNGTPATKREKDELTMWGEKANILVELISVGTGD---QMEPKK--TVT 163
DB 348 NVKFKLSKKGDSYVKNQDKKEIITDANGINNIK-----ALPSGDYILKEIAPRPYFD 402
OY 164 KAAEAOEIAK-TDKQ--FGT--EAAKVESAKAVTTQKV-----EYK----- 199
DB 403 KDKETPTMKDTONOGYFTTIEENAKAIEKTKDVSAOKWEGTQKVPITYFKLYKQDDNQ 462
OY 200 ---FSKA-VEKLTKEKEDKVT--NKAND-----KYVREVTLSSEDKRSATVELY---- 242
DB 463 NTTVDNAEIKKLEDGTTKVTWMLPENDKNGKAIKYLVKNVNAQGE--DTTPEGYTRKE 520
OY 243 -----SNEAAQOTYTVDNKVGK--TEVAVGSLS---EAKTIEMADQ--- 279
DB 521 NGLVVNTKERTPIETTSISGEKVMDDKDNQDKRPEKVSYNLANGEKKTLDVTSETNMK 580
OY 280 -----VVADEPTLOFTVDE-----NGTEVVSPECIEVTVAEKINAKGEITL 324
DB 581 YEFKDLPRYDEGKKIETVYEDHVKDYTTDINGTITNKK-----YTP-----GE-- 624
OY 325 AKGTSTTVKAVNY-----KKDKVNAESK-----EYKVSAGAFAVASINMT-----VAEON 370
DB 625 ---TSATVTKNMDDNNQDKRPEIEIVELYQDGAATGATKATAILNESNNWHTWGLDLSKA 681
OY 371 KADFTSKDFQNNKVEYEGDNAY--VOVELKQDFNAVTGKV--EYESINTEVAVVDKATGK 427
DB 682 KGOOVKTYVEELTKV---KGYTHVDNNDMGNLITNKKTPETTSISGEKVMDDKDN-- 735
OY 428 VTVLSAGKAPVYTVK--DSGKALVSHTVLEI EAPAKAMADI-----KLEKTNVALST 479
DB 736 ---ODGKREKVSYNLADGEEKVKTLDVSETNMKYEFDLPKYDEGKKIETVTEHDV 791
OY 480 KDVTDLKVKAPVLDOYGEKFTAPVTVKVLVLD---KDGKELKEOKLE---AKVYNRELVL 531
DB 792 KDVTDINGTITNKKYTPGETSATVTKNMDDNNQDKRPEIKVELYQDGAATGATKATAIL 851
OY 532 NAAQGEAGNTVVLTAKSGEKAKATLALAKAPGAFSKF-----EVRGLDTELD--- 581
DB 852 N---ESNNMTHTWT-----GLDEKAKGOQVKTYVEELTKVKGYTHVDNNDM 895
OY 582 -----KYVTE-----ENO--KNAMTVSVLPVDANGVLVKCAEAALKYTT 619
DB 896 GNLITVNNKTPETTSISGEKVMDDKDNQDKRPEKVSV---NLLANEKVKTTLDVTS 949
OY 620 -TN-----KEGKEVDAT-----DAQVTVQNNSVI 642
DB 950 EFMWKEFDLPKYDEGKKIETVTEHDVHYKDYTTDINGTITNKKYTPGETSATVTKNMDD 1009
OY 643 TVGQGKAGETKYKVVVLDGKLI-----TTHSEKVVDTAPATKALAVEFTSTSL 691
DB 1010 NNNQDKRPEIKVELYQDGAATGATKATAILNESNNWHTWGLD--EKAKGOQVKTYVDEL 1067
OY 692 KEY 694
DB 1068 TKV 1070

```

RESULT 10
SPAP_STRMU STANDARD: PRT: 1561 AA.
AC P23504:
DT 01-NOV-1991 (Rel. 20, Created)

```

DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DB CELT SURFACE ANTIGEN I/II PRECURSOR.
OS SPAP.
OC Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [11]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NG5 SEROTYPE C;
RX MEDLINE=91207143; PubMed=1982405;
RA Kelly C., Evans P., Bergmeier L., Lee S.F., Progulskie-Fox A.,
RA Harris A.C., Aitken A., Bleiweis A.S., Lehner T.;
RT "Sequence analysis of the cloned streptococcal cell surface antigen
RT I/II."
RL FEBS Lett. 258:127-132(1989).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=NG5 SEROTYPE C;
RX MEDLINE=91207143; PubMed=1982405;
RA Kelly C., Evans P., Ma J.K.C., Bergmeier L.A., Taylor W., Brady L.J.,
RA Lee S.F., Bleiweis A.S., Lehner T.;
RT "Sequencing and characterization of the 185 kDa cell surface antigen
RT of Streptococcus mutans."
RL Arch. Oral Biol. 35:335-385(1990).
RN [13]
RP -1- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: BELONGS TO THE SPAP/SPF/SPAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17390; CAA35253.1; -.
DR PIR: S06839; S06839.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.
FT SIGNAL 1 38
FT CHAIN 39 7
FT CHAIN 997 1561
FT DOMAIN 39 1536
FT TRANSMEM 1537 1561
FT DOMAIN 1557 1561
FT DOMAIN 1557 1561
FT DOMAIN 60 550
FT DOMAIN 219 464
FT DOMAIN 847 963
FT DOMAIN 1528 1533
FT SEQUENCE 1561 AA; 170060 MW; 540D92768FC8AB4B CRC64;

```

Query Match 5.4%; Score 211; DB 1; Length 1561;
Best Local Similarity 22.2%; Pred. No. 0.19;
Matches 181; Conservative 100; Mismatches 302; Indels 232; Gaps 39;

```

OY 25 GAVKND-----KGFEPGKELTRAEATMAQILNP-IDKDAKPSADSGOWYTP 76
DB 53 GTGTGNPATNLPRAQSGSAQADQSQTLEKRWVHTLEVKTLDLDAKAKAKSAG----- 107
OY 77 FIAAVERKAGVYKGTNGEPNGKIDRVSMASLLVEAVKLDTRVNGTPATKFKDLETLNWG 136
DB 108 -VAVVQDADADVNNKGTVTAE-----EAVQKETELEKEDYTKQAEIDIKK----- 147
OY 137 KEKANILVELGISVGTGDQMEPKRYTVKAEAAOIFAK---TDKQFG-----TEAKVE 186

```



```

OY 119 -----VNGTPATK-----KOLETLNMGKREKANIIVEIGI-----SVGTGDOMERK 160
DB 201 GTDIRHEITKETLTKLTKVTRME---WAOEGNSEDDLPJLTNPALIGL----- 251
OY 161 TVTREAEOFIKTDKOFTEAAKVESAKAVTTOKEVEKFSKAVEKLEKEDIKVTNNANN 220
DB 252 --IKANEVTLNGK-DAGIGNTYKV--ADGINANDFOQHQVMIKDKDEVIYMESEST 306
OY 221 DKVL-----VKEVITS-----EDKRSATYELYSNL-----AAKOTYTVDNKVKTEVA--- 264
DB 307 DQEVIMRVGEFTLKGTFEDPKDLSNDLADLELDAESEKSYRFENN---TKVTYNF 362
OY 265 -----VGSLEAKTIEMAD-----QTVVADEPALOP--TVKENGSEVVSPEIEVVT 310
DB 363 TRENDPVDGLKEITKIDNADSGFTGAKVYVLDNNNEIATIHVIDQSMNK--EDEGVKYG 420
OY 311 PAAEKINA-KGEIT-----LAKESTSTYKAVYK 337
DB 421 EVISKIDTDKKKITRNDKFNLDGKEGKDFVFLNGKPAKFSDLKEG---MYVSYY 477
OY 338 KDKVVAESKEVKVSAAGAAVASISNMTVAQNKADFTSKFKONKYYEG-----DNA 391
DB 478 ADGD---EDKLIVFATDTVVEGKVDK--VVSRRNNDYRLTIGDITYRYEGATESDGNK 532
OY 392 YVOVELKDOENAVTTGKVEYESLNTG--VAAYDKATGKVTLSAGKAPKVTVDKSKKAL 450
DB 533 DVODIDKDHMDLV-----DSLDETVKLYLDASGRVHTE-----TKDAIDRRKKAI 580
OY 451 VSHIVLEAFAQKAMKDIKLEKTVALSTKDVYDLKAKAPVLDQYGEFTAPVTVK-VLD 509
DB 581 VTSATFVN-----TSKDTWDFR---VLTQKKEITVTSLEAKNIYD 617
OY 510 KDCG-----ELKEOKLEAKYVNR---ELVNLAAQOAGVYVTLTKSSEKEKA 556
DB 618 FDKGNFSROKNODDLEDIIVPSKDKDTLLEVTYLDADG-KPGVVEELKPYKVQESKA 676
OY 557 TLALLEKAPAFSKFEVYR-----GLDTELDKYVTEENOKNAMTVSVLPVDANGLVLGA 610
DB 677 WDLADEDDDMVGSEVYTDKTAVERNMTGLEESSKREKLNKAKFKDY-----A 727
OY 611 EAELKVTNTNKEGKEVDADAQVTVQNNSVIT---VCG---GAAGEYKVTYVLDG 662
DB 728 DENDLSVITYVNDKDEVA---IFVEGDLGDAHYGVYIDGRRGKGTIRVWERKG 783
OY 663 KLITTHSFVY-----VDTAPFAKGLAVEFTSTSLKEV-----APNADLKALLNLS 709
DB 784 DKYVEKEYKLDGDDDLKDDDIRNDPIAFTVDSNDEVYVDVVEYVNNKAKHGLAEVTD 843
OY 710 VDOY-PATTAKTATASNEFVSADTNVVAENGTVGAKGATSIYKNLTVVKD 759
DB 844 EKGMKDANIDKMYVGLVSDYRKDT-IYYKDADNKKKAS---IKSATVYFD 890

RESULT 12
PAC_STRMU STANDARD; PRT; 1565 AA.
AC P11657;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, last sequence update)
DT 01-OCT-1994 (Rel. 30, last annotation update)
DE PAC PROTEIN PRECURSOR.
GN PAC.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TextID=1309;
RN 111
RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-48.
RC STRAIN=MT8148 SEROTYPE C;
RX MEDLINE=89343654; PubMed=2761390;
RA Okahashi N., Sasakawa C., Yoshikawa M., Hamada S., Koga T.;
RT "Molecular characterization of a surface protein antigen gene from

```

```

RT serotype C Streptococcus mutans, implicated in dental caries.";
RL Mol. Microbiol. 3:673-678(1989).
CC -1- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -1- SIMILARITY: STRONG TO AGGLUTININ RECEPTOR OF S.SANGUIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL: X14490; CAA32652.1; -.
CC PIR: S04729; S04729.
CC InterPro: IPR001899; Gram_pos_anchor.
CC Pfam: PF00746; Gram_pos_anchor.1.
CC PROSITE: PS00343; GRAM_POS_ANCHORING.1.
CC Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.
CC SIGNAL 1 38
CC CHAIN 39 1565
CC DOMAIN 39 1542 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 1543 1560 MEMBRANE ANCHOR.
CC DOMAIN 1561 1565 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 219 464 3 X TANDEN REPEATS, ALA-RICH.
CC FT DOMAIN 1532 1537 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
CC FT PROTEINS.
CC SO SEQUENCE 1565 AA; 170781 MW; 4C3B05C809D0C32A CRC64;

Query Match 5.2%; Score 201; DB 1; Length 1565;
Best Local Similarity 21.8%; Pred. No. 0.46; Mismatches 226; Gaps 37;
Matches 177; Conservative 97; Indels 226; Gaps 37;

OY 25 GAVRGND-----KGMEPEKELTRAPATMAQIILNLP-IDKPAKSPFASQOQWYTP 76
DB 53 GTORGNPATNLPEAGSGSKREABOSQTLEROMVHTIIVPTDLDQAQKAKKSAG----- 107
OY 77 FIAAVERAGVITKGTNGEFPNGKIDRVSMASLVEAYKLTDRVNGTPATKREDETLMWG 136
DB 108 -VNVAQADADVKNKGTVPKE-----EAVQKEIEIKEDYTKQAEDIKK----- 147
OY 137 KERANIIVELGISGTGQWEPKKTIVTKAEAOFIKTDKQFGTEAAKVESAKAVTQKV 196
DB 148 -----TTDQYKSDVAHAEVAKIKAK-----NOATREOYEKQMAAHKA 186
OY 197 EVKFSKAVEKLTG-----EDIKVTNNKAN--NDKVLVKKEVLTSEDRSATVELYS 243
DB 187 EVERINMANAMASKTAYEAKLIAQYADLAAYQKTAANAOAKALAAVQAELKRVQE--A 244
OY 244 NLAQKOTY--TVDNKVKGTEVAVGSLEAKTIEMADQTVVADEPAL--QFTVKDENGTEV 300
DB 245 NAAAKAAYDTAAVANNANKNEIIAANEE---IRRRNATAKAEYETKLAQYQAELEKRVQEA 301
OY 301 VSPGIEFVTPPAEKINA-KGEITFLAKGSTVTVKAVYKKQSKVVAESKEVVSAGAAVA 359
DB 302 NAANEADY---QAKLTAYQTELARVOKANADAKATV-----EAAVAANNAKNA 346
OY 360 SISNMVVAEQNKADFTSKDKQNNKVEGDNAYQVELKQDFNAVTTGKVEYE---SLN 415
DB 347 ALVAENAIKIQRNENAKATYEALIKQVEADLAIVK-----KNAANEADYQAKKLTAQ 399
OY 416 TEVAAVDKATGKVTVLSAGAPVK--VTYKDSKALVSHTEIEA--FAQKAMKDIKLE 471
DB 400 TELARVOKAN-----ADAKAAVEAAVAANNAALTAENTAIIKKRNADAKADYEAKLA 453
OY 472 K--TNVALSTKDVYDLKVPVLDQYKTEFTAPVTYVVDLQKDELKEQLEAKYVREL 529
DB 454 KYQADLAKYOKDLADY-----PVKIKAYEDEQSTIKALAE----- 489

```

QY 530 VLNAGQAGNTVTLAKSGEK-----EAKATLAE-----LKAPG-----AFSKFEVRL 576
DB 490 -LEKHEKNEGN-----LTPSAQNLVYLDLEPNANLSTTDGKFLKAAVDAFSK----- 538
QY 577 DTELDKYVEENQKNAMTVSLPVDANGVLKGAEEAEELKVTTNKEGKVDATDAQVTV 636
DB 539 STSKAKY-----DOK-----ILOLD--DLDITLLEQSNVYASSMELKNGFDAGKSTTV 586
QY 637 QNN-----VITVGQAKAGETKYTVVLDKLTJTHSFK-VVDTAPTAKLAV-----E 685
DB 587 SNNSQVKGMSVLLERQOSATATYTNLQNSYNGKISKIYKYTVDPKSKFGQKXVWLGI 646
QY 686 FTSTSLKEVAPNADLKALLNLISVDCVPATTKAKATASNVEFVSADTNNVAENGTVGAKG 745
DB 647 FTDPTL-----Gv-----FASAVTGOYEKN----- 666
QY 746 ATSIYVKNLTYYKDGKEQKVEFDKAVOVAESI 777
DB 667 -TSIFIKNEFTFHEDEKPIFNFNALLSTVSL 697

RESULT 13

GUN_BAC66 STANDARD: PRT: 941 AA.
AC P19424;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (ALKALINE CELLULOSE).
OS Bacillus sp. (strain KSM-635).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1415;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91037937; PubMed=230718;
RA Ozaki K., Shikata S., Kawai S., Ito S., Okamoto K.:
RT "Molecular cloning and nucleotide sequence of a gene for alkaline
RT cellulase from Bacillus sp. KSM-635.";
RL J. Gen. Microbiol. 136:1327-1334(1990).
CC -1 CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1 SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -1 SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M72420; AAA2304.1; -
DR PIR: S29043; S29043.
DR InterPro: IPR001347; Glyco_hydro_F5.
DR Pfam: PF00150; cellulase. 1.
DR Pfam: PF00395; SLH. 3.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5. 1.
DR PROSITE: PS01072; SLH_DOMAIN. 2.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
FT SIGNAL 1 29
FT CHAIN 30 941 ENDOGLUCANASE.
FT DOMAIN 40 99 SLH 1.
FT DOMAIN 100 151 SLH 2.
FT DOMAIN 152 225 SLH 3.
FT ACT_SITE 373 485 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 485 485 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 941 AA; 104628 MW; BEA2AC3B169BPADA CRC64;

Query Match 5.1%; Score 199; DB 1; Length 941;
Best Local Similarity 20.2%; Pred. No. 0.31;
Matches 179; Conservative 126; Mismatches 343; Indels 238; Gaps 42;

QY 5 FPDVPADHHGIDSINVLVEKANKGNDKGMFEFGKRLTRAEATMAQILNLPIDDAKP 64
DB 41 FSDVKRTSWEFYIKDLYDEYVITGTSATTFSPDSTVTRQFTVTLRGIGLEASSKDPY 100
QY 65 SFADSGOWTPPTIAVERKAGVIGKTGNG-FEPNGKIDRSMASLLVEAYK-LDTKVN-G 121
DB 101 -FKDRK-NMAKYKIQAAIYERGIYTGKTNGEFAPNENITRQMAAAMVRAIYIENELSLP 158
QY 122 TPATKFKDELTLNW-----GKEKANILVELGISVGTGDOMEPPKTVTKABAQFI----- 171
DB 159 EEOREYNDSSSISTFAQDAVOKAYVL-EL-MEGNTDGYFQPKRSTREQSAKYISTLWK 216
QY 172 -AKTDKQFGEAKKAVSAAVTTQKVEV--KSKAYEKLKEIK--VINKANNDVLYK 226
DB 217 VASHDYLYHTEAVKSPS-EAGALQVELNGQLTLAEGDGTVPQLRGMSTHGLQWFEIYN 275
QY 227 E--VTISEDKRSATVELYSNLAAKQTYTVDNKVG-KTEVAVGSLEAKTIEMA---DOT 279
DB 276 ENAFVALSNDKGSNMTRL-----AMYGNGYATNPYKDLVYGLIELAFEHDMY 325
QY 280 VVAD-----EPTALOFT-----V 292
DB 326 VIVDMVHVAQGRADVYSGAYDFEEIADHYKDHKNHYIWELANEPSNNNGGPGLT 385
QY 293 KDNENGEVVSPEIEFVTYPAEKINKGETTLAKGSTTYKAYKKDKGVYASKEVYKS 352
DB 386 NDEKGEAYK---EYAEPLVEMLREKGDMLIVGN----- 417
QY 353 AEGAAVASISNMTVAEONRAD--FTSKDFKONKKVEGDNAVQV-----ELKQFNAV 404
DB 418 -----PMSQRPDLSDNPIDANENIMYSVHFTGSGASHIGYPSGPESSKSNM 468
QY 405 TTGKVEYESINTEVAAYVDKATGVYVLSAGKAPVYKVDKSKALVSHVTEIAPAQRA 464
DB 469 --ANVRV-ALDNGVAVF--ATEWGTQANGDGGPYDEADVWMLFNKHTISWANNSTLN 523
QY 465 MKDIKLEKTNVALSTKDVDYDLKAVPVLQYKEFPYAVYVYKLDGKELKQKLEAKY 524
DB 524 KNEISGATFPFELGRTDAD-----LDPGANQVAP-----BELSLSGEY 563
QY 525 VNRVELVLNAGQAGNTVTLVTLAKSGEKEKATLALAKPAFSKFEVRLGJDTLTKYV 584
DB 564 VRARI-----KGIEYTPI-----DRTKFKLVWDFNDGTGQFQVNG-DSPKKEI 608
QY 585 TEENQKNAMTVSLPVDANGVLKGA--EAEELKVTTNKEGKVDATDAQVTV-----Q 637
DB 609 TLSNNNDALQIEILVNSND--ISEGYNWNVRLSADGWSNVDILGATELTIDIVIEEPT 666
QY 638 NNSVITVGGQAKA-----GETYKRVYVLDKLTJTHSFKVVDPTAP 678
DB 667 TVSIAAIPOGPAAGMANPTRAIVTDEDPESFGDKALKAVTTISE--DSLSLETIATSP 724
QY 679 AKGLA--VEFTSTSLKEVAPNADLKALLNLISVDCVPATTKAKATASNVEFVSADTNNVA 736
DB 725 DNTMSNLIIFVGT-----DAD-----VISLDNI--TVSGTEIEIIVHDE----- 763
QY 737 ENSTVAKKATSIYVKNLTYYKDGKEQKVEF--DKAVOVAVSIKEA 780
DB 764 -----KGTATL-----PSTFEDGTGQGMWHTESGVKTALTEEA 798

RESULT 14
NCAL_CHICK STANDARD: PRT: 1091 AA.
ID NCAL_CHICK
AC P13590; Q90919; Q90918;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1999 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)


```

FT DOMAIN 624 692 FIBRONECTIN TYPE-III 2.
FT DOMAIN 152 156 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 161 165 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 41 96 BY SIMILARITY.
FT DISULFID 139 189 BY SIMILARITY.
FT DISULFID 225 287 BY SIMILARITY.
FT DISULFID 329 385 BY SIMILARITY.
FT DISULFID 426 479 BY SIMILARITY.
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .).
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .).
FT VARSPLIC 702 726 STPTSGTAAIVGILVIFVLL -> TLGSPSTSSFV
FT VARSPLIC 727 1091 SLLSAVTLILLC (IN 726 AA ISOFORM).
FT VARSPLIC 771 809 MISSING (IN 1052 AA ISOFORM).
FT VARSPLIC 810 1070 MISSING (IN ISOFORM N-CAM 140).
SQ SEQUENCE 1091 AA; 117415 MW; B151367002DF88E0 CRC64;

```

Query Match Best Local Similarity 19.3%; Score 193.5; DB 1; Length 1091; Matches 158; Conservative 136; Mismatches 313; Indels 211; Gaps 33;

```

OY 17 SINYLVKGVKGVNDKGMFE-PGKELTRAEATMAQ-ILNLPIDKAPSPFADSGQWY 74
DB 170 SNNYLQIRG-IKKTDGTCRGCRILARGEINPKDIQIVNPPSVARAGSTMA----- 223
OY 75 TFFIAVEAGVYKGTGKNF-EPNGKIDRVSMASLLVEYKIDTKVNGIPATKFKLETL 133
DB 224 ---TANLSQSVTLACDADPEPT-----MTWTKDGPEDNEEKY 263
OY 134 NNGKREANLIVELGISVGDGMEPKVTAKAOFIAKTOKFSTEAKVESAKAVTT 193
DB 264 SNNYDSEILIT-----KAYDKSDEATYICIAEKKAGEQATIH-LKVPK 307
OY 194 QREVKFSAVERLTKEDIKYTNKANDRVLYEVTLSEDKRSATVELYSNLAKOTYV 253
DB 308 PRITYVENKTAMEL-EDQTLTCEASGDP-----PSITKSTSRINSEKTL 355
OY 254 DVNKKYKTEVANGSEAKTIEAD-----QIVVADEPT---ALQTVKDEKTEV 301
DB 356 DQRIYVSHARVSLTLKLEIQTADAGEVYCTASNTIGOSQAMYLEVOYAPKIQGVAVY 415
OY 302 SPECIEFYVPAEKINAKETLAKGTSTYKAVYKKDKVVAESKEVYSAAGAVASI 361
DB 416 TWEG-----NQVN-----ITCEVFAVYSAVSWFRDQLLPSSNYSNIKIYNTPSASY 463
OY 362 SNNYVAEONKADFTSKDFKONKKYEGDNAYVOVELKDQFNNAVTTGKVEESINTEVAV 421
DB 464 LEVTPRESEN-DFGNYNCTAVNRIGESSEFIIQ-ADTPSPSIDRVEPYSSTARVERD 520
OY 422 D-KATGKIVVL-----SAGKAPVAVYVYKDSKGLVSHYVIEFAQKAMDITLEKT 473
DB 521 EPEATGAGVILYKKAEMWRLGEGEWSRLYDAK-EANVSGTTISGLKDETTYVRLS-- 577
OY 474 NVALLSTKDVTDKVKAPVLDQYKTEFAPVYVLDKDEKLEKQLEAKYVRELVLNA 533
DB 578 --AVNKKGGEISLPSDFKTPYREPSAKPLGEOGEDNSIK-----VAV 621
OY 534 AGQEAQ-----NYTVVLTAKSGEKEKATLLELKAPGAFSKEVGLD--TELKRYTE 586
DB 622 IKODDQSGPIRHYLIKRYKAKH--SEWKP--EIRLPSGIDHMLKSLDMNAEYEVYIA 676
OY 587 ENQ-----KNAMTVSLPVUDAN----- 603
DB 677 ENQQGSKPAHYAFRTSAQTYIAPASTSPTSGITAAIVGILVIFVLLVAVDTCYEL 736
OY 604 ---GLVL-----KGAEAELEKVTITNKEKE--VDATDAOVYONNVITIV 644
DB 737 NCGILMLCAIVNLCKGSGFGAKGDKMEBGKAAFSKDESKPEIYVTEEBERPNMD---G 793

```

```

OY 645 GOGANAGETIKYVT-----VVLGDKLITTHSFKYVDAPATKGLAVEFT 687
DB 794 GKHTPEENETPTLPEPEHTADTAATVEDMLPSVTTG---TTNSDTITETPATQONSPST 850
OY 688 STSLKEVAPNADLKAALLNLSVDCVPATTKATASNV 725
DB 851 TLTSTSIAPPA---AIPDSNAMSPOQATPARAKASPV 885

```

```

RESULT 15
MRP_STRSU STANDARD; PRT; 1256 AA.
ID MRP_STRSU
AC P32653;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE MURAMIDASE-RELEASED PROTEIN PRECURSOR (136 KDA SURFACE PROTEIN).
GN MRP.
OS Streptococcus suis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1307;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SEROTYPE 2 / D282;
RX MEDLINE=92267650; PubMed=1587602;
RA Smith H.E., Vecht U., Gielkens A.L., Smits M.A.;
RT "Cloning and nucleotide sequence of the gene encoding the
RT 136-kilodalton surface protein (muramidase-released protein) of
RT Streptococcus suis type 2."
RL Infect. Immun. 60:2361-2367(1992).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, X64450; CAA45781.1; -.
CC DR PIR, A43829; A43829.
CC DR InterPro: IPR001899; Gram_pos_anchor.
CC DR Pfam: PF00746; Gram_pos_anchor. 1.
CC DR PROSITE, PS00343; GRAM_POS_ANCHORING. 1.
CC KW Repeat; Transmembrane; Cell wall; Signal.
CC FT SIGNAL 1 47
CC FT CHAIN 48 1256
CC FT DOMAIN 48 1231
CC FT TRANSMEM 1232 1248
CC FT DOMAIN 1249 1256
CC FT REPEAT 663 681
CC FT REPEAT 839 861
CC FT REPEAT 953 1006
CC FT REPEAT 1064 1084
CC FT REPEAT 1089 1142
CC FT REPEAT 1143 1195
CC FT DOMAIN 1223 1228
CC FT PROTEINS.
SQ SEQUENCE 1256 AA; 135795 MW; DCF7F65242F14341 CRC64;

```

Query Match Best Local Similarity 4.8%; Score 187.5; DB 1; Length 1256; Matches 195; Conservative 114; Mismatches 299; Indels 403; Gaps 47;

```

OY 31 DGMFEPGKELTRAEATMAQILNLPIDKAPSPFADSGQWYTPFFIAVEAGVYKGT 90
DB 299 EKYTIEPGEVSVFPHSTKYNANNSNITYTYDTS-----LASANTPGALAKS 344

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2002, 07:51:27 ; Search time 111.59 Seconds
(without alignments)
1028.979 Million cell updates/sec

Title: US-09-754-947-1
Perfect score: 3885
Sequence: 1 AGKTFPDVPADHWGIDISIN.....EPDKAVQAVSIEKAPATK 785

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.17 : *
1: sp_archaea : *
2: sp_bacteria : *
3: sp_fungi : *
4: sp_human : *
5: sp_invertebrate : *
6: sp_mammal : *
7: sp_mhc : *
8: sp_organelle : *
9: sp_phage : *
10: sp_plant : *
11: sp_rodent : *
12: sp_virus : *
13: sp_vertebrate : *
14: sp_unclassified : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3304.5	85.1	816	2	Q9ZES5 Q9ZES5 bacillus th
2	770.5	19.8	821	2	Q9RED0 Q9RED0 bacillus th
3	748.5	19.3	823	2	Q45759 Q45759 bacillus th
4	719	18.5	862	2	P94217 P94217 bacillus an
5	487.5	12.5	531	2	Q9RMZ0 Q9RMZ0 bacillus an
6	449.5	11.6	931	2	Q91655 Q91655 bacillus fi
7	445	11.5	1099	2	Q68840 Q68840 bacillus st
8	415	10.7	920	2	Q45664 Q45664 bacillus st
9	398.5	10.3	1047	2	Q9RB35 Q9RB35 cytophaga s
10	348.5	9.0	1268	2	Q9REB7 Q9REB7 bacillus sp
11	279.5	7.2	404	2	Q9X324 Q9X324 bacillus an
12	268.5	6.9	969	5	Q9NDP9 Q9NDP9 plasmodium
13	268	6.9	1036	2	Q9B999 Q9B999 clostridium
14	259	6.7	1109	2	Q53505 Q53505 campylobact
15	249.5	6.4	3198	5	Q9UB88 Q9UB88 manduca sex
16	248.5	6.4	1112	2	Q52781 Q52781 campylobact
17	245.5	6.3	1822	2	Q07290 Q07290 streptococc
18	241	6.2	1128	5	Q26947 Q26947 trypanosoma
19	240.5	6.2	2353	2	P71401 P71401 haemophilus

20	240	6.2	3488	5	P91257 P91257 caenorhabdi
21	237.5	6.1	652	2	Q9X360 Q9X360 bacillus an
22	235.5	6.1	1344	2	Q49545 Q49545 mycoplasma
23	234.5	6.0	1795	2	Q9LCJ9 Q9LCJ9 staphylococ
24	234.5	6.0	6642	5	Q01761 Q01761 caenorhabdi
25	232.5	6.0	1557	2	Q9RN12 Q9RN12 haemophilus
26	232.5	6.0	1758	2	Q9JMS5 Q9JMS5 escherichia
27	232.5	6.0	1829	5	Q22248 Q22248 caenorhabdi
28	231	5.9	903	2	Q9K105 Q9K105 bacillus st
29	231	5.9	1432	2	Q9F292 Q9F292 yersinia pe
30	231	5.9	2349	2	P94750 P94750 escherichia
31	231	5.9	2383	2	P76347 P76347 escherichia
32	230	5.9	2402	2	Q9AER7 Q9AER7 staphylococ
33	229	5.9	6632	5	Q17362 Q17362 caenorhabdi
34	228.5	5.9	2055	2	Q85472 Q85472 abiotrophia
35	227.5	5.9	5458	5	Q9U459 Q9U459 plasmodium
36	226	5.8	1097	2	Q9ZAI7 Q9ZAI7 anaerobic t
37	225.5	5.8	2045	2	Q9AOK5 Q9AOK5 streptococc
38	223.5	5.8	3241	2	Q9AHR9 Q9AHR9 fusobacteri
39	222	5.7	565	2	Q30921 Q30921 mycoplasma
40	221.5	5.7	6677	5	Q9N435 Q9N435 caenorhabdi
41	220.5	5.7	1148	2	Q60045 Q60045 thermoaer
42	220.5	5.7	2478	2	Q9RL69 Q9RL69 staphylococ
43	220	5.7	1752	2	Q9A552 Q9A552 rumiococcu
44	218.5	5.6	1302	2	Q49547 Q49547 mycoplasma
45	218.5	5.6	1323	2	Q9AL34 Q9AL34 clostridium

ALIGNMENTS

RESULT	ID	Q9ZES5	PRELIMINARY:	PRT:	816 AA.
AC	Q9ZES5	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	CTC PROTEIN.				
GN	CTC.				
OS	Bacillus thuringiensis.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OX	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID=1428;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CTC;				
RA	Sun M., Yu Z.;				
RT	"S-layer protein gene of Bacillus thuringiensis CTC."				
RL	Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AJ012290; CAA09981.1; -				
DR	InterPro: IPR001119; SLH.				
DR	InterPro: IPR003343; Big_2.				
DR	Pfam; PF00395; SLH; 3.				
KW	Pfam; PF02368; Big_2; 1.				
KW	S-layer.				
SO	SEQUENCE	816 AA; 87293 MW; 1002892905802151 CRC64;			
Query Match	Best Local Similarity	85.1%; Score 3304.5; DB 2; Length 816;			
Matches	679; Conservative	33; Mismatches	68; Indels	5; Gaps	5;
Qy	1	AGKTFPDVPADHWGIDISINYLVEKAVGNDKGMPEPKELTRAAATMMAOILNLPDK 60			
Db	30	AGKTFPDVPADHWGIDISINYLVEKAVTGNDKGMPEPKELTRAAATMMAOILNLPDK 89			
Qy	61	DAKPSFADSGOGWYPTFLAAVEKAGVIGKTGNGFEPNGKIDRVSA\$LLVEAYKLDTRVN 120			
Db	90	DAKPSFADSGOGWYPTFLAAVEKAGVIGKTGNGFEPNGKIDRVSA\$LLVEAYKLDTRVN 149			
Qy	121	GTPAKRFKDELTMNGKRAKNTLVELGTSVG-TGQWMPKRTVTA\$AQAQFIAKIDKQFG 179			

```
Db 150 GTPATPKFQDLETLNMGKEKANILVELGISVGTADKWEPRKVTYKAQAQFIKTDKQFG 209
Qy 180 TEAAKYESAKAVTTQKVEVFKSAVEKLTREKEDIKVTNNKANNDKVLKREYLTSEDKRSATY 239
Db 210 TEAAKYESAKAVTTQKVEVFKSAVEKLTREKEDIKVTNNKANNDKVLKREYLTSEDKRSATY 269
Qy 240 ELYSNLAQKOTYTVDNKVGKTEVAVGSLEAKTIEMADQTVVADEPTALQFTVKGDENGTE 299
Db 270 ELYSNLAQKOTYTVDNKVGKTEVAVGSLEAKTIEMADQTVVADEPTALQFTVKGDENGTE 329
Qy 300 VVSPEIEIEVTPAAEKINAKGEITTLAKGSTYTVKAYKKKGDKGVAAASKVEKVASAEGAANA 359
Db 330 VVSPEIEIEVTPAAEKINAKGEITTLAKGSTYTVKAYKKKGDKGVAAASKVEKVASAEGAANA 389
Qy 360 SISNMVTVAEONKADFTSKDFKONNKVYEGDNNAVQVELKDOFNAVTTG-KVEYESLNTVEY 418
Db 390 SISNMVTVAE-KADFTSKDFKONNKVYEGDNNAVQVELKDOFNNVNNKAEYESLNTVEY 448
Qy 419 AVVDKATGKVTYVLSAGKAPVYTVKDSKGKALVSHTEIEPAQAQKAMKDIKLEKTVNAVS 478
Db 449 AVVDKATGKVTYVLSAGKAPVYTVKDSKGKALVSHTEIEPAQAQKAMKDIKLEKTVNAVS 508
Qy 479 TKDVTDLKVKAPVLDQYGEKFTAPVTVKVLVDKDKELKQKLEAKTVNRELVINAQOEA 538
Db 509 TKDVTDLKVKAPVLDQYGEKFTAPVTVKVLVDKDKELKQKLEAKTVNRELVINAQOEA 568
Qy 539 GNTVTVLTAKSCEKAKATLAELEKAPAFSKEVEKGLDTELDRKYVTEENOKNMTVSVL 598
Db 569 GKTVTELTAKSCEKAKATLAELEKAPAFSKEVEKGLDTELDRKYVTEENOKNMTVSVL 628
Qy 599 PVDANGLVLKGAFAELKATTTTKEKGEVDATDAQVTVQNNNS-VTVGOGAKAGETTYKYT 657
Db 629 PVDANGLVLKGAFAELKATTTTKEKGEVDATDAQVTVQNNNS-VTVGOGAKAGETTYKYT 688
Qy 658 VLDGKLITTHSHKVVDTAPTAAGLAVETSTLKEVAPNADLKAALNLINLSDGVPAAT 717
Db 689 VLDGKLITTHSHKVVDTAPTAAGLAVETSTLKEVAPNADLKAALNLINLSDGVPAAT 748
Qy 718 AKATASNEVEVSADTNVA-ENGTVGAKGATSIYKVLTVVKGDKQKQVEFDKAVOAVS 776
Db 749 AGATVTDKVFVSADTNVSEETAKFGTKGSTIFVKELFVKKGCEQOKVELDKPRVADVS 808
Qy 777 IREAK 781
Db 809 IREK 813

RESULT 2
Q9REDO PRELIMINARY: PRT: 821 AA.
AC Q9REDO: 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE SURFACE-LAYER PROTEIN PRECURSOR.
CN SLPA.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 4045;
RA Mesnage S., Hausant M., Gounon P., Fouet A.;
RT "Characterization of B. thuringiensis S-Layer.";
RL Submitted (Sep-1999) to the EMBL/Genbank/DDbJ databases.
DR EMBL: AJ249446; CAB63252.1;
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 3.
KW Signal.
FT SIGNAL. 1 29
SQ SEQUENCE 821 AA; 87279 MW; 8D68995C812214B8 CRC64;
```

```
Query Match 19.8%; Score 770.5; DB 2; Length 821;
Best Local Similarity 32.5%; Pred. No. 2.8e-26;
Matches 283; Conservative 114; Mismatches 305; Indels 169; Gaps 40;

Qy 1 AGTFFPDVADHWGIDISINLYEKGAVKNDKQMEPEKELTRAEAATMAOILNIPDK 60
Db 30 AGTFFPDVADHWGIDISINLYEKGAVKNDKQMEPEKELTRAEAATMAOILNIPDK 89
Qy 61 DAKPSFADSGOGWYPFIAVEKAGYIKGTGNG-PEPNKIDRVSAVSLVEAYKIDTV 119
Db 90 GAKPSFKDADSDMAAKYIAVEQAGVIGDGTGNEPNSQININASHASITVAKLIEDKY 149
Qy 120 NGTPATPKFQDLETLNMGKEKANILVELGISVGTQDQMEPRKVTYKAQAQFIKTDKQFG 179
Db 150 SGLELINFQDLKD-HMGEKANILVALGITNGNGMDEPDKSTRBAKFIKTDQMGFG 208
Qy 180 TEA-AKYESAKAVTTQKVEVFKSAVEKLTREKEDIKVTNNKANNDKVLKREYLTSEDKRSAT 238
Db 209 OKAEKAVESIKEINAKGEITTLAKGSTYTVKAYKKKGDKGVAAASKVEKVASAEGAANA 266
Qy 239 VELYSNLAQKOTYTVDNKVGKTEVAVGSLEAKTIEMADQTVVADE-PTALQFTVKGDE 295
Db 267 ITLKNALVKNQAVAHVKD-----VKSVDGKIDIFKALEVIFPNDVAPVTVSTVSTPBG 319
Qy 296 NGTEVVSPE-----GIEFVTPAAEKINAKGEITTLAKGSTYTV----- 332
Db 320 NVKVFSEKLSKDAVTVVINGKEFTATPEE-----NTVTLRADVASVKNGEAFNVIVT 374
Qy 333 -----KAYKKKQDKVYA-ESKEKVS--AEGAANVSSISMTVAEONKADPF 374
Db 375 GAKDLVGNTEMVEGKATYKVEKDVTAPEVKDKIKVELDVGATVLEEV-----TSEELSAO- 430
Qy 375 TSKEFKONNKVYEG-----DNAYVOV--ELKDOFNAVTTGKVEY-----ESLNTVEYA 419
Db 431 GKTVYKKGKGVVAGATIALDQDTKALVTPPALITKEKTAANL-KVEFYGYDAANN---- 486
Qy 420 VVDKATGKVTYVLSAGKAPVYTVKDSKGKALVSHTEIEPAQAQKAMKDIKLEKTVNAVS 479
Db 487 VGNKVTKEKVTKDVVAPNVLKVVADENKA-ATFTFDKQVTAQDEG-----KLRYINLDS- 540
Qy 480 KDVTDLKVKAPVLDQYGEKFTAPVTVKVLVDKDKELKQKLEKTVNRELVINAQOEA 539
Db 541 KDVTKEVAVASVED-----NKKAITLTFQEGNKKVAA--TKGFVADTPAGNES 587
Qy 540 NYTP-----VVLTAKSCEKAKATLAELE-KAPGAFSKEVEKGLDTELDRKYVTEENOKNMT 594
Db 588 AFTKEVYKVEKKEGKKDEKAPATKVERVADSKTFEY-----TFDKVY--KGGGADS 640
Qy 595 VSVLPVDANGVLKGAFAELKATTTTKEKGEVDATDAQVTVQNNNSVTVGOGAKAGETTYKYT 654
Db 641 AS-----NVANNYTLAGAKLEBGLIIVNADGK-----SVTIELPETFTF---EKSETV 685
Qy 655 KVTYV-----LDG-KLITTHSHF-KVVDTP-APTAAGLAVETSTLKEV-----APADK 701
Db 686 KETVANAVANKDQYKMGITMLVNVVDTKAPBER-SAKITKYDAKIIITLTFSEAVAVD-- 741
Qy 702 AALLNLISVDGVPATYAKA--TASNVEFY-----SADTNVAENSTGAKGATSIYKVN 753
Db 742 -ATDFEYIDNGVVLPTATKADETAESKDVVLKVTAPADVNLATGTVTKAVEKGVVLN 800
Qy 754 LTVVKGDKQKQVEFDKAVOAVSIAKAPAT 784
Db 801 TADTSTINKNKLAVF-----KPV 818

RESULT 3
AC Q45759 PRELIMINARY: PRT: 823 AA.
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
```

DE CRYSTAL PROTEIN.
GN CRYTKD.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TKD2-14;
RA Hashimoto N.;
RT "Bacillus thuringiensis subsp. mexicanensis TKD2-14,"
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases
DR EMBL; D86346; BA013073.1; -.
DR InterPro; IPR001119; SLH.
DR Pfam; PF00395; SLH; 3.
SQ SEQUENCE 823 AA; 87602 MW; 96FA83B36A8BACE7 CRC64;

Query Match	19.3%;	Score 748.5;	DB 2;	Length 823;
Best Local Similarity	31.5%;	Pred. No. 2.6e-25;		
Matches 277;	Conservative 116;	Mismatches 302;	Indels 185;	Gaps 41

QY	1	AGKFEPPADPHMIDISINTVEGAKVAGKNDKGFEEGKELTTRAEATMAOJLINTPDK	60
Db	30	AKGSEFPDVPQBSMSAEIYDILYAKKALIEGKPDGTFAPTEAIDRASAKMAITLLEGEKE	89
QY	61	DAKPSFADSOQOMYTPFIAAEKAGVIRKGTGNG--FEENGKIDRYSMASSLVEAYKIDKTV	119
Db	90	GAKPSFKDAQOSMAKAYIAAEKAGVITOGDEBQCFENFNNOINRASMATMIVKAYKIDGKV	149
QY	120	NGTPATKFKDLETLNMGKEKANILVELGISVGTCDONEPKYTYTAKAEOFIKTDKQFG	179
Db	150	SGOETKPSDJKD--HMGKKAANIILVALDITNGCTGNGEPPKSVTRAEMAEFIKTDQFG	208
QY	180	TEA-AKVESAKAVTOKVEYFSAVEKLTRKEDIKYNKANNNKVLKEXYTLSEDRSA-	237
Db	209	QKAEKVESIKAINAKKELVAKLGTVTEBYEYKNNKFFLLSGVDVATYVADKIGAKSVI	268
QY	238	-TVELYSMLAKOTYTVDVNKGTEVAVGSLEAKETEMADQTVADDEPALOFTVADEN	296
Db	269	LTLEDGRKLEKNMAYVYTVKE-----LMTAEKELPKALVETFEFDEVA-----PT	314
QY	297	GTEVVSPEGIEFVTPAAEKINAKETILANGSTYTVKAYIKKDGKVAESKEVKSAGEA	356
Db	315	VSTVSTPBG-VKVKVFSSEKIDAAKTIV-----VYNQCKVEGKEVNN	355
QY	357	AVASISMTVAEONKADPT-----SKPEKQNN-KYVEGDMAYYVELKODFNMTTGKVE	410
Db	356	TFT--SNQALMLQNGKEPSIIVTGATDLADMTMEYEGKATY-----KVE	398
QY	411	YESLNTVEAVADKATGKVTYLSA-GKAPVKATYVD-----SKKALVSHTEI	457
Db	399	KDTIAPKEVKD-----KYKETTAGSKATLEVTSEBLATNGQGVYVKKKEELKSTIAY	453
QY	458	EAPQOKAMKDIKLEKTYNALSTKDVTDKY-----KAPVLDQYGEKFTAPV	503
Db	454	D--AODTKTAAV-IEVTGALKKEKATDITVEEYVYKDAANNVSGSKVAKVSKDVAPN	510
QY	504	TVKV-----LDKD-----GKLEKQDKLE-AKYVNRBLVNLNAGQDAGNYTVLT	546
Db	511	FVKVEADONKATFTPEKVEYTAONGK--LRVINLDTSDVDYATVAPAKNNKGITLTFP	569
QY	547	AKSGEKEAKATLALTELKAPG---AFSKFEYRGIDTELDRKYVTEENOKNMYFVSJLVPD-	601
Db	570	AKGNYKVA-AKKGLVLDAGNESALFK-EKKYVEKEPAEK-----ETDKVAPVYKAVAYDK	623
QY	602	ANGLY-----LKGAELAEIKVTTN-----KEGKEDATDAQVTVONNSVITVGQ	646
Db	624	ATNKVTVWFEKDEYKGGQVAESAAVANNVYTLAGAKLPGEGTLIVDGTKAITELSTFTF--	681
QY	647	GAKGGEYKYVYVY-----LDGKLITHS--FRYVDT-APYAKGLAVETYSISKEV-----	694
Db	682	-EKSEYVTKFVAVANKDGKMMGTANLLEVDKTPKPEEK--SAKITYDAEAEITLTFES	737

```

0y 695 -APNADLKAAALLIISVDGVGATPAKATAS-----VEFSAPTINVAEGVGAAGAT 747
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 738 EAVAVNDITDFVLD---NGVELYIAKKADATFEAAKDVVLKATATDVLNANGITYYK--- 792
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0y 748 SIYKYLITLVADGKEQYEF---DKAOVAVSIEAKPAT 784
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 ---AKELE-----NNAKYEFTTKIENALAF---RPVT 820
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

P94217	
ID	P94217
	PRELIMINARY;
	PRT;
	862 AA

```
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE EAG GENE PRECURSOR.
```

0C	Bacillus/Staphylococcus group; Bacillus.
0X	NCBI_TaxID=1392;

RP SEQUENCE FROM N.A.
RC SPRAIN-STERNE DERIVATIVE:
RX MEDLINE=97260111; Pubmed=9106206;
RA Wessene S., Jost-Culture E., Mook
RT "Molecular characterization of the Bacillus anthracis main S-layer
RL component: evidence that it is the major cell-associated antigen." J
RT Mol. Microbiol. 23:1147-1155(1997).

DR	InterPro: IP0001899; Gram_pos_anchor.
DR	Pfam: PF00395; SH3, 3. Gram_pos_ANCHORING; UNKNOWN_1.
DR	ProSite: PS00343; GRM_POS_ANCHORING; UNKNOWN_1.
KW	Signal; S-layer.
FT	Signal 1..29 POTENTIAL.
FT	CHAIN 30
SEQUENCE	862 AA: 91362 MW: Cbl6B202F62CCAC CRC64

Query Match	18.5%;	Score 719;	DB 2;	Length 862;
Best Local Similarity	29.6%;	Pred. No. 5.4e-24;		
Matches 233;	Conservative 123;	Mismatches 289;	Indels 142;	Gaps 29;

[illegible]

QY 175 DKQFGTEAKVE-----SAKAV-----TTQKV----- 196
Db 211 E-----VEEERIRPGLVPTVSSVTAVNATTLTVLSDTLTVEVTLDTALVANEATEVTFE 266
QY 197 --EVEFSKAVEKLTREDIKVTKANKANDKVLKVEVTLSEDDKRSATVELSNLAAKQTYTND 254
Db 267 INDVEYETLVTVYEBELAAVANAAGKOLVNFOTOPVADASSVLTPASTASTVTKGVLS 326
QY 255 VNRKV-----KTEVAAGSLEAKTIEMADQTVVADEPTALO-----TVKDENGTEVVSPE 304
Db 327 LNRSTDRHSNBNVGLVAERAVLSSDGKLTJVTAPTGQFFKNQNDVTVADASGQNTIPS 386
QY 305 GIEFVT-----PAEKI-----NAKEI----- 322
Db 387 VIKTISVDOTTAPIVNSVYVVPATDKFEVTLSEPIDLSLGEVLRINGOPVASGFDALTCP 446
QY 323 -----TLAKGSTT----- 331
Db 447 TNKLTFFARPSSVATGTNATYTAGFSDAAGNFVTPSTTVPVTOOTTAALAVASLEQVTNQ 506
QY 332 --VKAVYKKDKGVVAESKEKVSAAEGAAY-----ASISMTVAEONKADFTSK--DFKQNN 383
Db 507 KRLTFENKE--LNSASKTALLESCTGVVYTRPNSGTSNTFTVAQNLTVDNKGTYDITLSD 564
QY 384 KYEBGDNAVY--QVELKQDFNAVTTGKVEYESLNTBEAVAVDKATGVTVLSAGKAPVKYT 441
Db 565 ATYTGNSSEVFCITLTKDAFTDVTNNK--NDLYSKSITLKNDTVAFTVTSALASNRQA 621
QY 442 VDSKGAALVSHVTEIEAFQAKAMDKLEKTNTVALSTKDVYDLKAVKAVLDQYGEFPA 501
Db 622 I-----EYTLSCGVITTNPAQVKLRKDKAEQGTGLSVALKGGTD----- 659
QY 502 PVTYKVL-D-KDGKELK-----EOKLEA-----KYVNRLEY-----LNAQO 536
Db 660 --NVLVSVSSGAELAGSGYQVLEAGVATDINGNANNANVAPSVSATPAPLNVAYA 717
QY 537 EAGNTTVVILTAESGEKAKATLALBKAGASKEFEVRDLDELKYTYEENQNAAYVS 596
Db 718 NSGTFTVTAFTAGQTFETVYASL-----NHNNEFKIDQAVSSNSDITLNRITRITVS 769
QY 597 VLPVDA-----NGLVLRKGAABELKVTNNKEGKEVDADAQVTVONNSVITVG 645
Db 770 LPSEDSVKISGALFTTNGLAL-----ESKRALATATATVYTDN----- 809
QY 646 OGAKAGETTKVTVLDGLKITTHSEKVVDTAPATAGLAVEFTSTSLKEVAPNADIKALL 705
Db 810 -----TAPPLNGAQL-----VSANYI 825
QY 706 NILSDGVPATTAKATASNEFEVSADTNVVAENGTVGAKGATSIYKNTLYVK--DGK 761
Db 826 KTFEDENIDLALTDAADLIDIDLOLNGTVAYNGGF--ASGATS--GGDVTVVSVDCK 879

RESULT 7
O68840 PRELIMINARY: PRT: 1099 AA.
AC O68840:
01-AUG-1998 (Tremblrel. 07, Created)
01-AUG-1998 (Tremblrel. 07, Last sequence update)
01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SUPSPACE LAYER PROTEIN.
GN SBSC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
CX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC12980;
RA Jarosch M., Egelseer E.M., Mattanovich D., Sleytr U.B., Sara M.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF055578; AAC12757.1; -
InterPro: IPR003343; Big_2.

DR Pfam: PF02368; Big_2; 3.
SQ SEQUENCE 1099 AA; 115394 MW; 7330F156334ABBA CRC64;
Query Match 11.5%; Score 445; DB 2; Length 1099;
Best Local Similarity 24.6%; Pred. No. 8.1e-12;
Matches 223; Conservative 136; Mismatches 331; Indels 218; Gaps 41;
QY 44 AEAATMAAOLN-----LPIDKDKPSFADSGQWYTFPIAV 81
Db 28 SQAATDVAIVTVSQAQAKQKKAAYTTSHTVTETGCPDPDKDYAAATNKKK--DANANAVAV 86
QY 82 EKAGVIKGTG-----NGEPPNGKIDRVSNASLLEAVYKIDTVNG-----TPAT 125
Db 87 NKAGGAKKDAYLADQAIYETVYFKANPKSGEARVAT--YIDAYNATKLRDKMROELKAAV 145
QY 126 KFKDLETLN-----WGKE-----KANIIVELGIS 149
Db 146 DAKDLKKAEBELYHKISTYELKTRTVILDRVYGOSTRELLRSTFFKADAQALRRLIYDITVA 205
QY 150 VGTGDQWEPKK--TVYKAAEA-----QPIAKTDKQGTG-----AAKVE 186
Db 206 MKARAQDAVAKAGNLDKAKKAAALDQVNOYVSKYTDFAKRELQKAQDAKATAEALTPKVE 265
QY 187 SAKAVTTQKVEYKESKAVEKLT--KEDIKVTNKANDKVLKVEVTLSDNKSATVELYSN 244
Db 266 VSAIDSTSFKYTEFKPYDKATAIPIKNFSITLKGTEFKLPRKSVSESGLATATVLYDT 325
QY 245 LAAKQTVVDVKKVKTVEVAAGSLEAKTIEMADQTVVADEPTALO----- 290
Db 326 LYDGKTYTVVTS--GLKTPAGKEFETSTNEFTYKPV--PASITFNKMLPESAVULT 380
QY 291 --TVKDENGTEVVSPEGIEFTVPAER-----INAKGEITL-----AKGTS--TVKAV 335
Db 381 KVVTVKDAAGNVTKS--GPELEFTSSEKLTQCKPINTGKKSIVYANATVKGNTVTTGVI 438
QY 336 YKKDKGVVAESKEKVASBGAAVASISMTVAEQKADFTSKDFQONNKYVEGDNAVYQV 395
Db 439 IAVEDEKAAEVELKLTDKNKEVTL-----YANGNAFD-----EDGQOISSG--TLTLTA 487
QY 396 ELKQDFNAVTTGKV-----EYESLNTBEAVAVDKATGVTVLSAGKAPKTVYKDSKRAL 450
Db 488 KFKDQYGMELTGKAGTDTYTESLNPVLYV--APGCVTPYIPGALYKVKYGE----- 540
QY 451 VSHVTEIEAFQAKAMDKLEKTNTVALSTKDVYDLKVRAPVLDQYKGEFTAPVTVKLDK 510
Db 541 VTKTIPVTVKAMPVLEJTAVDSTGVSAKGQKATPKV--TLKDQYGNKFTGVNV----- 593
QY 511 DGELEKQKLEAKYNNRELVLNAAQOEAGNTV-----VITAKSGEKEAKATIAL 560
Db 594 -----TSDKTETATVS-----VSNSGIGOSEYTVTVNGVAEGSTTITISGTEVEK--VPV 642
QY 561 ELKAPGAFSKPEVRGLDT--ELDKYVTEENOKNAMVSVLPVDANGLVKGAEAALKVT 619
Db 643 NVVAGGPVANTQIKVLDGKIDKSKATESPANNQVLKAYAVDANGNIY--GDTINDVTTIS 701
QY 620 --TNKEGKEVDATDAQVTVONNSVITVGOAKAGETTKVTVVLDGKLITTHSEKVVDTAP 677
Db 702 EATDINGVIYVANSKSTANGDPTVYITDNGSKVKGE--TLTVKLGTVLTGTVDAVEYIDT-- 758
QY 678 TAKGLAVEFTSTSLK--EVAAPNADLKAALLNIIISV--DGVPAITTKATASNE----- 726
Db 759 TLKATVTVTKKADLIELDADNGDALAKLLANLIDKQGNPVDVSAAPNTNNEKLOALKS 818
QY 727 ---FVSADTVNVAENGTV--GAKGATSIYKVLTVVKQKGEKV-----EFGKAVOVAVSI 777
Db 819 VLSGIVSSDTVIGSVSVNDNLKDASI--SGLAVKKGAGTVTLTVFNEDSKIAPIALTV 876
QY 778 KEAKPATK 785
Db 877 K-APATQ 883

```

RESULT      8
ID 045664    PRELIMINARY:      PRT:      920 AA.
AC 045664:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE SBSB GENE (SEQUENCE 5 FROM PATENT WO9906567 PRECURSOR).
GN SBSB.
OS Bacillus steatothermophilus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PV72;
RA MEDLINE=97197531; PubMed=9045827;
RA Kuen B., Koch A., Asepauner E., Sara M., Lubitz W.;
RT Molecular characterization of the Bacillus steatothermophilus PV72 S-
RT layer gene sbsB induced by oxidative stress.;
RL J. Bacteriol. 179:1664-1670(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PV72;
RA Lubitz W., Resch S.;
RT "SECRETION OF CARRIER-BONDED PROTEINS INTO THE PERIPLASMA AND THE
RT EXTRACELLULAR SPACE.";
RL Patent number WO9906567-A/5, 11-FEB-1999. LUBITZ WERNER (AT);
RL RESCH STEPHANIE (DE).
DR EMBL: X98095; CAA66724.1; -.
DR EMBL: AX000222; CAB77070.1; -.
DR InterPro: IPR003343; B1g_2.
DR InterPro: IPR001119; SLH.
DR Pfam: PF02368; B1g_2; 2.
DR Pfam: PF00395; SLH; 3.
DR S-layer; Signal.
KW CHAIN
FT SEQUENCE      920 AA; 97916 MW; 1F3C20344B40F3A2 CRC64;

```

```

Query Match      10.7%; Score 415; DB 2; Length 920;
Best Local Similarity 24.4%; Pred. No. 1.3e-10;
Matches 207; Conservative 123; Mismatches 352; Indels 168; Gaps 35;

QY 4 TFPDVPADHMGIDSINYLVEKGAVKGNCKGMPEPKELTRAEATMAQILPLDIDKAK 63
DB 33 SFTDVAPOY--KDAIDELVSTGATKGTETKPCVDEITRLDAAYLAVLVDVDMKD 90
QY 64 PSFADSGOGMYTFPIAAVEKAGYIKGTGNG-FEPNCKIDRVSMASLVEAYKLD----- 116
DB 91 AGTTPKPD-RAKYVALAEAGVLNKGAKPGFAYDPLTRVEMAKTIARRYKLKADVKL 149
QY 117 --TRVNGTPATKFKDELTLMCKEKANILVELGISVGTGDQDQEPKITYTKAEAFIKT 174
DB 150 PFTDVNDTWAPYKALYKEVTKGK-----TPTSAGAYONTIRGDFQAFVRA 197
QY 175 DMOFGTEAKVSAKAVYTKQVEKFKSAVERKLTREKIDKVTNKKANDKVLVEKVTLSKD 234
DB 198 VNI--NAVPEIETVAVNSTYKVFNTQI-----ADVDFNFDLNGLTJTKATLSDK 250
QY 235 RSATVELYSNLAKQTYTVDVNKG--KTEVA-----VGSF-EAKTIEADQTVVADE 284
DB 251 KSEVEVVVKPFTRNQDITITATGINKLGEYAKELTGKFWVSVDADYVALNNSLSKGE 310
QY 285 PIALQFTYKDENGTEVY-----SPECIEFTVPAEKINAKGEITLAKGTSTTVKAY 336
DB 311 EESGL--TVKDDGDKVDYAKVELTSSNTNIYVSSGEVSAAKAVTAVKPTADVTAKVT 368
QY 337 KKDGVVAES-----KEYKVASGAAVAASISNMTVAEO-----NKADFTSKDFKONK-- 384
DB 369 LPDGVVLNTRKVVYTVPEVPOVONOGFTLVDNLSNAPONTVAFAKAKAVYSMEFGETKTV 428
QY 385 -VYE--GDNAVVOVELKDPQFNAVTTGKVEYESLNTETVAVAVDKATGCVTVLS-----AGKA 436

```

```

DB 429 AMYTKNGDETRKPVDEKD-----ATVRSLNIIATAINSEELVYTNAGOSGKA 479
QY 437 PVKTVKDSGKALVSHITVEIEFAOKAMKDIKLEKTNALS-----TRQVTDLKVK 488
DB 480 SFEVTEKDN-----TKRTEFVDKDPVLODIDKVDATSVKLSDEAVGGEVGVNQKTIK 534
QY 489 APYLDQYKEF-----TAPVTVKVDKQ-----GKEIKQ----- 518
DB 535 VSAVDQGEIKETKGTGKVTITNTGVLKYNVNSNTIDPFGSASADQEVVAVATKDKI 594
QY 519 ---KLEAKYVNRVLVNAAGQEAQNTVVLAKSGEKAKAT-LALELKAPGAFKFEVR 574
DB 595 VNGKEVKYFKN-----ASDTPTSTKTIIVNVAVKADATPGLDIAP--SEIDVN 645
QY 575 GLDTELDKRYTEENOKNATVSVLPDANGVLKGEAEALKVTTNKEGKEVADTDQV 634
DB 646 APPT--ASTADVDFINFSVELTYLDSNGNKL-----KVPT--ATLVGTNDYV 692
QY 635 TYQNNSVITVGGG-----AKAGETVYVTVLDGLITTH-SFRVVDPTAPKGLAVEFTST 689
DB 693 EVNGNVLPQKGNDELTLTSSSTVNVVDVADG--ITKRIPVKYINSASVPASATVATSPV 750
QY 690 SEKEVAPNADLKAALNLTLSVDGVPATATKATASVEEVSADTNVVAEMGVAKGATSI 749
DB 751 TVKLNSDDDLTFEELIGVID--PTQLVKDEIDN-ETIAV-----SKAAKNDGYL 798
QY 750 YKKNLTIVKD 759
DB 799 YNKPVLTVTKD 808

RESULT      9
ID 09RB35    PRELIMINARY:      PRT:      1047 AA.
AC 09RB35:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE S-LAYER PROTEIN PRECURSOR.
OS Cytophaga sp. 'Jeang 1995'.
OC Bacteria; CF8 group; Cytophagales; Cytophagaceae; Cytophaga.
OX NCBI_TaxID=85173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JEANG 1995;
RX MEDLINE=95291197; PubMed=7539663;
RA Jeang C.L., Lee Y.H., Chang L.W.;
RT "Purification and characterization of a raw-starch digesting amylase
RT from a soil bacterium--Cytophaga sp.";
RL Biochem. Mol. Biol. Int. 35:549-557(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JEANG 1995;
RA Jeang C.L., Liao T.W., Chlou S.Y., Kang P.L., Shieh T.Y.;
RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF068060; AAF21259.1; -.
DR InterPro: IPR000515; BPD_transp.
DR InterPro: IPR001119; SLH.
DR InterPro: IPR003343; B1g_2.
DR Pfam: PF00395; SLH; 3.
DR Pfam: PF02368; B1g_2; 2.
DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBER; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL      1 32 POTENTIAL.
FT CHAIN      33 1047 S-LAYER PROTEIN.
FT SEQUENCE   1047 AA; 108718 MW; 4395402E9965295A CRC64;

```

```

Query Match      10.3%; Score 398.5; DB 2; Length 1047;
Best Local Similarity 24.2%; Pred. No. 8.4e-10;
Matches 228; Conservative 111; Mismatches 365; Indels 237; Gaps 40;

```

[illegible]

```

Oy 589 QKNMNTVSVLPVDANGVLYLKAKEAAELKVTTTNECK-----EVDATDAO-VTV 636
Db 704 -----VGTGDI-----ANAFALITNSGKTLTNLIGKMTYELRNTSGOPIEV 745
Oy 637 -----QNNSVITVVGOGAKAGETVAVTVLDOGLKLTTHSFKVDPDAPRAKLAVEFTS 688
Db 746 KTPISYRQNNVSSLT---FRIEYGSITTSAD-VISTSNAAPLIEAPRKGKATVYVP 801
Oy 689 TSLKEVAPNADLKA-ALLNIIISVDC-----VPATTA-----KATASN-----VEEVA 750
Db 802 SFV--TANNFPASADNNINGVVTVDARDPNNKHTVPALPATAFNGNKKEVGTNYTVOGVFNK 859
Oy 731 DTNVVAENGST-----VAKAKATSIY 750
Db 860 DKKTTFNGKNAVYVAGEGKTYRY 884

```

RESULT	ID	PRELIMINARY:	PRT:	404 AA.
09X324	09X324			
AC	09X324:			
DT	01-NOV-1999 (TREMBlrel. 12, Created)			
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	PX01-54.			
OS	Bacillus anthracis.			
OC	Plasmid virulence plasmid PX01.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1392;			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-STERNE.			
RA	Kochina R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Keim P.			
RA	Kochler T., Lamte G., Kumano S., Mahillon J., Manter D., Martinez Y.			
RA	Ricke D.O., Svensson R., Jackson P.J.;			
RT	"The sequence and organization of pX01, the large Bacillus anthracis			
RT	plasmid harboring the Anthrax toxin genes.",			
RL	J. Bacteriol. 0:0-0(1999).			
DR	EMBL; AF065404; AAD3358.1; -			
DR	InterPro: IPR001119; SLH.			
DR	Pfam: PF00395; SLH: 3.			
DR	Prosite; PS01072; SLH_DOMAIN; UNKNOWN_1.			
DR	Plasmid.			
QO	SEQUENCE	404 AA;	45044 MW;	45D08FAA550C0C4C CRC64;

[illegible]

```

Db      313  G--MYLPFGSLKAQFIYIPINSGSFKVIYFLDNNDATAYELAKWTTLKSLDNLDEIQGT 370
QY      344  AESKEVVKTSAGCAVAASISNNTVAEÖKADPTSKDFQÖNNKVVYEGDAAYÖVEIK 398
        11 :: -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
Db      371  VESHKLNTFEFGGNKVKVRIGNST-----ADNM-----MYÖVESK 404

```

RESULT 12		
C9ND19		
ID	C9ND19	PRELIMINARY; PRT; 969 AA.
AC	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE	MEROZOITE SURFACE PROTEIN 3G.	
GN	MSP-3G.	
OS	Plasmodium vivax (strain Belm).	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=31273;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=BELEM;	
RA	Gainsel M.R., Al-Khedery B., Ingravallo P., Corredor-Medina C.,	
RA	Barnewell J.W.;	
RT	"Plasmodium vivax merozoite surface proteins-3b and -3g share	
RT	structural similarities with Plasmodium vivax merozoite surface	
RT	protein-3a and define a new gene family in Plasmodium.",	
RL	Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.	
SO	EMBL; AF099663; AAF8288.1;	
SO	SEQUENCE 969 AA; 104674 MW; 9A150CCA9918FF77 CRC64;	

Query Match	Similarity	6.9%	Score 268.5	DB 5	Length 969
Best Local	Similarity 22.8%	Pred 0.00038			
Matches 185	Conservative 102	Mismatches 335	Indels 191	Gaps 32	
QY	23	EKGAYKGNDRKGFEEGKELTRAEEA	----	TJMAQILNLPIDDAKPSFASOGOWYTPET	78
DB	110	EEDAKAKQK - ILEAVDRYKKAEEAIKKGTKAEAIKNAVNK - AK - EVADAKAKELTT	----		163
QY	79	AAVEAGAIVIKGTGNGFEPNGKIDRYSMSASLVEAYKKLDTKVNGTATYKFKDLETLNMGKE	138		
DB	164	--EKAKIAKEAEETAKIGKSVSASFALLIAKEAKETAKE	-----	ANE	204
QY	139	KANILIVEIGISVGDDOMEPPKTTVAKAEAOPIAK	-----	TDKQFGTFA	182
DB	205	KQAMAEALATHY - RNAYEAKKEAEKAOVANAARKEIEMLSKITYKEEKADTTEESATYA	262		
QY	183	AK-----VESAKAVTTQREYV-----RFSKAVEKLTK - EDIKVTNKANNDKVLVEYT	229		
DB	263	AESASQVSEYKAKGEVGAKEAKAALNLTDAVERLEKESSELLKDNVLRDTVNSLKEGA	322		
QY	230	LSEDRSATVELYSNLAKQTYTVDNKNYGVKTEVA-----VSLSEAKTITEMAQTYVADBP	285		
DB	333	TEEQKRAKKEEKAKIS-----EEYAKKEAASAKOFATIEERANYEANKIAENHP	373		
QY	286	TALQFTVDENQETEVSPGIEFTVPAA - EKNAKGEITIAKGTSTYKAVYKKGKVYA	344		
DB	374	NT - NVTEEANKAKASTKASTEATKASTEATNAASTEATKPSKANKYK-----KKDEAK	428		
QY	345	ESKEVYVSAEGAAYVASISNMTVAEQNKADFTSKDEKQNNKYVEGDNAYVQVELKDQFNAY	404		
DB	429	AAKEAKKAKTEAYIALPVT KAMAAREKAKKSAE-----AADKAKQAQAEAVNGASER	479		
QY	405	TTGKAYEESLNTNEVAVDKATGYVLSGK--AAVYKTVKRSKRGALVSHIVLEAFQ	462		
DB	480	TKKDAEHATKAN-----EKKHTETADAANKNAEKVVEEDNVAKN-----EKKK	527		
QY	463	KAMKRIKTEKTVALTSTRDVTDKAPVLDQYKGEFPAPYVYK - LDKQKELKEOKL	520		
DB	528	KKVDVY - IEKVLEAKSEE-----DTYQAOIQAEIIVQAVANVEACEKAKTAQ	575		
QY	521	EAKYVNRRELVINAQOEGNVTVYLTKASGEKEAKATIALELKABGAFSKPEVRGLDTL	580		

Db 576 EKKAKDEAV--KAKEAE-----AKKQAEKREKIT----- 605
Oy 581 DKYVEENOKNMTVSVLPDANGVLKGAEAELKVTTNKEGKEDATDAQVTONNS 640
Db 606 -KTAEE-----ANKAKEEAKSSEAKQAEFTAGD-----VDEEVAVNNE 646
Oy 641 VTTVGCGAGAGETKYVTVLDGKLITTHSFKVVDPAPTAKGLAVEFTSTSLKEVAPNADL 700
Db 647 FESVAAAAAHHKVPETLD-----KEKKNAENAKKASAKATEAKTTAETATKATEA 701
Oy 701 KALLNLTISVDGPATTKATASNVEFSADTNVAENGTGCA-----RGATSIYVK 752
Db 702 KAA-----AGNACKASENKAIRAD--VLAEKASTEAKSLKEAKKLADIKKS 748
Oy 753 NLTVYKDGKQKVEFDKAVQAVASIKAKPA 783
Db 749 NVTNEKAKRDKAANDAAHQASLSASKAKEA 779

RESULT 13
086999 PRELIMINARY; PRT: 1036 AA.
AC 086999;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE S-LAYER PROTEIN.
GN SLPA.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10682;
RX MEDLINE=98129094; PubMed=9467913;
RA Lemaire M., Miras I., Gounon P., Beguin P.;
RT "Identification of a region responsible for binding to the cell wall
within the S-layer protein of Clostridium thermocellum."
RT Microbiology 144:0-0(0).
RL EMBL: U79117; AAC33404.1;
DR InterPro: IPR001119; SLH.
DR Pfam: PF00395; SLH; 1.
SQ SEQUENCE 1036 AA; 113329 MW; 1AF2202060D7C2A1 CRC64;

Query Match 6.9%; Score 268; DB 2; Length 1036;
Best Local Similarity 20.5%; Pred. No. 0.00044;
Matches 217; Conservative 133; Mismatches 325; Indels 382; Gaps 50;

Oy 16 DSINY-----LVEKGAVGNDKGMPEPEKE--LTRAENATMAAOI-----LNLPIDK 60
Db 27 DSEFYEKEAEILYRLGLYKGTSETVEVPLNEGLDQKOTGVMLRLFGCEDDALLETPEME 86
Oy 61 DAK---PSTADSO--GQWTPPIA-AVEKAGVKTGTNG--FEBNGKIDRVNASLIVE-- 111
Db 87 AAOGLAAKFKDAADLADMAQROGVAAYVER-GYVKGYPDGTFLPNADLNGIAFCSLILOOL 145
Oy 112 -----AYKIDT--KVNCTPATRKDLETLN-----WGKEKA 140
Db 146 GYDGDVFDEDAAYKLOEFGGLTAEOAEAFNNKNGINRDSNVGIAFSALQAVYKATGTVI 205
Oy 141 NLIVELG-----ISVGTGQMEPEPKTVTKAEAAQFIKTDKQGTAAKVESAKAV----- 191
Db 206 EYLVENGANSKELATLELGLVLLAKIKVKKALDAVK-----QVGRPVPULPEEVEVYEDD 259
Oy 192 TTQKVEVKE-----SKAVEK--LTKEDIKVTN--RANND 221
Db 260 TTEKLAWEPTVDTSEVSGQETIGTIKASGLAYREPKATLKVIYPEELQVVDVAKPUL 319
Oy 222 KVLVE-----VTLSEDKRSATYELVLSNLAKQTYTYDVAK 257

Db 320 KEIVIEFNGEVAASKADEKSSVEDNTIELVYSEDKTTVTITLVAGAMTAEEIEVTI-- 377
Oy 258 VGKTEVAVGSLKAKTIEMADQTVVADEPTAL-----QETVK-----DENGTE 299
Db 378 --KTATGLKEEYTKFVVPADENPEAESIALIGPNSFEIKFSEPYOSSDAEVLVNDGYT 435
Oy 300 VVSPE-----GIEFVTPAEKINAKG-----BITLAKGS--TT 331
Db 436 YVSEELKSDYPTLVLVELGVSSINEGTYKVKGYRDVAGNIMRTKTEFLERVKDTPPT 495
Oy 332 VAAVYKKDKGVAAESKEVAV-----SAEGAFA-----SISNMTVAEQKADFTSKDFQ 381
Db 496 AAVKEATQNKVYIEFNEPATRGGYSGDEALATRDIFQYQYSSMKPKVYASD----- 547
Oy 382 NKKVY-----EGDNAYVVELKQDFNAVTGKVEYESLNTAEVAVDAKGVKTVYLSAG 434
Db 548 NKKVYTVLYSESDNDGY-----PVYLLPVG 573
Oy 435 KAPVYTVVDSKGLALYSTVELEAFQAKMDIKLEKTNVALSTKDYTDLYKAPVLDQ 494
Db 574 NVTITI-LKEVDDAVV-----DAGNKLSEDLKLTAT--VAADNAPTVKSVTAEDK 625
Oy 495 YKEFTAPVTVK-----VLDKGKELKEQKLEAKYVNRRELVNAAQHGAGNTVVL 546
Db 626 IYVFESEDVNEQAKDKMDYVLIKDKGEID-----TAISS 660
Oy 547 AKSGEKEKATATALELKADGAFSKEVREL-DTELD---KVT---EENOKNAMTVSVL 598
Db 661 IYDYSNETKVTIYLDKELSGKTYTIDIKGITSSENEKMAVTTIEFTVDTAPTEIEV 720
Oy 599 PDANGVLVKGAEAEELK--VTTNKEG--KEVDATQAOVTY-----QNNSVTVGQ 646
Db 721 TFDVNTIYVYSEAMSTKNGSVLNKDKYLVDDNKVKEIKKIELEFGDKMKNVRIYVS 780
Oy 647 GAKAGETVYKVVY-----VLGKLTITTHSF-----KVY 673
Db 781 DVDLNDVYELTIANVDEAGNNAISAFDVAKKAKLSEQAEVSEIRISTRIEIVINKLL 840
Oy 674 DPAAPAK-----GLAVEFTSTSLKEVAPNA-----DLKALLNLT 708
Db 841 DKATVEKTFEVEERGSKNVALTRISSITYDDGKTIYKGLPDAVRPANGDITGYLYI- 899
Oy 709 SVDGVAATTA--ATASNVEFSADTNVAENGTG--AKGATSIYVKNL----- 754
Db 900 -GELIKSDTKEMANGAVSKPVDKFPSPFSVANGVYGDASKGFTLFDIDIKFLNNS 958
Oy 755 -----TYVKDQK---EOKVEFDKAV---QVAVSIR 778
Db 959 AGLGATDVLVKKSGKTLLEAGIDYDAVAIDNKITVTLK 995

RESULT 14
053505 PRELIMINARY; PRT: 1109 AA.
AC 053505;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE TYPE A S-LAYER PROTEIN.
GN SARP2.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=95204338; PubMed=7896695;
RA Dworkin J., Tummuru M.K., Blaser M.J.;
RT "A lipopolysaccharide-binding domain of the Campylobacter fetus S-
layer protein resides within the conserved N terminus of a family of
silent and divergent homologs."
RT J. Bacteriol. 177:1734-1741(1995).
DR EMBL: S76860; AAB33871.1; -.


```

Db 1274 -----TVAEASTEPSSTF-----AITSUK----- 1292
QY 416 TEVAVVDKATGKVTVLKAGKAPKVTVDKSKGKALVSHTEIEAFQAKMDIKLEKTNV 475
Db 1293 TKMSTISEETGKTSVSE-----EITVKT-----TWTEESTEPSSTEAITSEKTSV 1337
QY 476 ALSTPDVT-----DLVKAPVLDQYKGFAPYTVKVLKDGKELKQKLEAKYVNR 528
Db 1338 -ISTSEETGKTSVSEEVTKTVTDE-ATEITSTVSI-----ETKE----- 1376
QY 529 IVLNAGOEAGNYTVLTAKSKEKAKATLLELKAPGAFSKFEVRLDTELDKVTEN 588
Db 1377 -----TSVSGSTELST-----QASSKIESPTTESGITSHTHESE 1411
QY 589 QKNAMTVSVLPVDANGLVL-KGAEAELKVTTNKEGKEVDATDAQVTVQNNSVITVGOG 647
Db 1412 ESTVSTTEKGEVTSETTELKTSIVSTEMLSTTEKE----- 1447
QY 648 AKAGETVKTVVLKGLKLTTHSFKVVDTAPPTAKGLAVEFTSLKEVAPNADLKALLNI 707
Db 1448 -----ETVWSTTELSTEEF-----SSTTESTETGSTE----- 1477
QY 708 LSVDCVPATTAKATASNTEFVSADTNVVAENCTVGAKGATSIYKKNLTVVKDKGEQKVEF 767
Db 1478 ISTETESTTAS-----QIIPASTTVAPETNETGTP-ASALPVEGLSTSTEG-----ITT 1526
QY 768 DKAVOVAVSIREAKPAT 784
Db 1527 QESEVTVTTEKSKRPGT 1543

```

Search completed: March 23, 2002, 08:01:55
 Job time: 628 sec

378 TTCTCAAGTAACCTTGGGCTTAATAAATCTAGTAATCGCAGTAGTGA 427
146 euGllyIleSerValGlyThr...GlyAspGlnThrPgluProIlyStrythr 161
428 ACGGCAATTTTCGAAGTACTGATGCAACAACTTAACCTTAACCACTCA 477
162 ValThrLysAlaGlnAlaGlnPheIleAlaLysThr...AspLysGln 177
478 ATCACTCGTCAGACCTTTCGACTAGTGTTCAAACGTCAGTGTGACAAAGT 527
177 nPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrG 194
528 TGAAGTCAAACTCCAGAAAGCAGCATTTGTTAAAGCTATCAACACA 577
194 LuIysValGluValLysPheSerLysAlaValGluLysLeuThrLys... 209
578 CAACCTGTGAAGTACATTGCAAGAGAAAGTAAGTACTAAGCTCAAGCATT 627
210GluAspIleLysValThrAsn.....LysAlaAs 219
628 AACTCAAAATCGAAGCTTTAGAAATTAATAATGCTTGTAAACCAAC 677
219 nAsnAspLysValLeuVal.....LysG 227
678 AAACAAAAAGTTGTTCTATTACTAGTCAAGCTCAACAGCTGATTAAG 727
227 LuValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSer 243
728 AGATGCTTTAACTCTTGAC.....GGCGAACAATCGGTGGCTTTAA 771
244 AsnLeuAlaAlaLysGlnThrTyrThrValAspValAsnLysValGlyLys 260
772 GGGTGGCTGCT.....GTAGTCCAACTAAAGTTGGAAC 806
260 sThrGluValAlaVal...GlySerLeuGluAlaLysThrIleGluMetA 276
807 AGATCTTGTGCGAGTTCAGGTAACCTTGTCAAGAAAGTAAGATTCAAG 856
276 IaAspGlnThrValValAlaAspGluProThrAla.....LeuGln 289
857 CTAAAGTAACTGTTGCTGAAGGTCATCTAAAGCTGATTCCTGTTACT 906
290 PheThrValLysAspGluAsnGlyThrGluValLysSer..... 302
907 TTCTACTGTACAGTAACAACAATGATGGCTGTATACCAACATTTAACAG 956
302 302
957 TGAAGCTTTAAACAAGCAAGAGGATCGCAACATCTTACACTCCT 1006
303 ..ProGluGlyIleGluPheValThrProAlaAla..... 313
1007 ATTAAGCAAGTACTGATGAAGTAACGTATTGCAACACTGTGATGCTTCT 1056
314GluLysIleAsnAlaLys 319
1057 AAATTTCTACTGGTTATGTATTTCTGGGTGATGATACCAATTCCTTCACT 1106
319 sGlyGluIleThr.....LeuAlaLysGlyThrSerThrThrV 332
1107 TGAAGCAAGTAACTAGTGTGCTTCAATTAATGATGTCGAACAAACATT 1156
332 alLysAlaValTyrLys.....LysAspGlyLysVal..... 342
1157 ACAAAAGTACTATTAATAAACCTTAATAACTGTAAACCGAAGCAAAACAA 1206
343ValAlaGluSerLysGln 348
1207 ACATTTAATGTTGGTTTGTAGAAAAATGATGTCTTACTTCTGATTAAGT 1256
348 uValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTrpT 365
1257 ACGAAATGCTACAGTTATGGCGTTAAAGCAATTAACAATTAAGCAATGTA 1306
365 hrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
1307 CAGCTTTAGAC.....GCTGCTCAAAATTAACAAGATTTCTAAAGGTGAA 1350
382 AsnAsnLysValTyrGluGlyAspAsnAlaTyrValGluValGluLeuLys 398
1351 GCTACATTACAGATTTCGTGACTACATATGACACTGTGAAGCCAGTAGTGA 1400
398 sAsp.....GlnPheAsnAla 403
1401 TGATCTACACAGCACTAACAATAGTACTTCAATAAATAATATAGTCAT 1450
404ValThrThrGlyLysValGluTyrGluSerLeuAsnThr 416
1451 CTGCTTTCAAACTGCTCTTCTAAAGTAACTTTCGCTGCTTCAAGCA 1500
417 GluValAlaVal.....ValAspLysAlaThrGlyLysValTh 429
1501 GAGTATACAAATTTGAGTTAACTACGTGCTGATTAATGCT...GGAGAAGTTGC 1547
429 rValLeuSerAlaGlyLysAla...ProValLysValThrValLysAsp 445
1548 TGCAATTTGGCCTACTAAGCGTCCGGAATTAACAAGTTATGTTAAAGATA 1597
445 eLysGlyLysAlaLeuValSerHisThrValGluIleGlu..... 458
1598 AAGCTGTAACTTAGCTAAATGAATGAACGTTAATGTTGCATTCAATGAA 1647
458 458
1648 GATAAAGTCGTGTAATTTCAACAGTTTACAATGCTTAATTCGTGATAC 1697
459AlaPheAlaGlnLysAlaMetLysAspIle 469
1698 TGATCCAAATCTCAGATATCTTACACAGGCGATTAAGCAAAACAATCT 1747
469 yLysLeuGlyLysThrAsn..... 474
1748 CTGTA...AAAACAATGATTAAGGTGAAGCTACATTTGTTATGCTTCT 1794
475ValAlaLeuSerThrLysAspValThrAs 484
1795 GATACAGTAACGATTAATGCAACACCAATTCCTTGATGATTAATATAC 1844
484 pLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGluPheThrA 501
1845 TTCTGATCAAAACAGCGACGCTTGAATGAAGGTGAACCAAAACAGATG 1894
501 IaProValThr...ValLysValLeuAspLysAspGlyLysGluLeuLys 516
1895 CACCAATCTTACTTCCAAACACCATATCTTGAATGGCTGAGTATGAAA 1944
517 GluGlnLysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAl 533
1945 GCATACAAA.....AAATCAGATCTTAATAAAGCTGAACATACTT 1985
533 alAlaGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLysSerG 550
1986 TGATGCTTCAAACTGACAGATTTTGAAGCAATTTAGTAACCAAAACGG 2035
550 LyGluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGly 566
2036 GCAAAAAGTACGTCGTAATCTTATTT.....AAGAAAGCA 2070
567 AlaPheSerLysPheGluValArgLysLeuAspThrGluLeuAspLysTyr 583
2071 ACTTATACAAATCTCAATACGTGCGCTCAATCATATTTAAAGTACATAACA 2120
583 rValThrGluGluAsnGluLysAsnAlaMetThrValSerValLeuProV 600
2121 AGTATATCCACCAATGCTAGCTACAGAGTAACTTATGAAGCT...ACTT 2167

[illegible]

Patent No. 5874267
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Expression of surface layer proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/682,517
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3666 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3666
US-08-682-517-14

alignment_scores:	Quality: 378.50	Length: 1009
	Ratio: 0.762	Gaps: 44
Percent Similarity:	49.257	Percent Identity: 22.101

```
alignment_block:
US-09-754-947-1 x US-08-682-517-14 ..
Align seg 1/1 to: US-08-682-517-14 from: 1 to: 3666
```

16	AspSerLeuSrnIryLeuValGluIyusGlyAlaValLysyGAsnAspIy	32
43	GAAGCACTTCAACCTTAACTTATTAACCAAGCGATATCCAAAGGTACTAA	92
32	sgIyMetPheGluProGlyLysGluLeuThrArgAlaGluAlaIatHm	49
93	CGGAACTTCAACCCACTTAAACAAGTAACTGCTGCACAACTGCAGAAA	142
49	etMetIaGlnIleIeuAsnLeuProIleAspLysAspAlaLysProSer	65
143	TCCTTCACAAAGCCTTTAGAGATTAGAAAGCTAACGGAGATGA.....AAC	186
66	PheIaIAsp...SerGlnIyLeuIrrPyrThrProPheIleAlaIaIaIa	81
187	TTCAAGACGTGAAGACGTGGCGCTGTGTACTACAACTCAATCGCTGCTGT	236
81	IgluIyAlaGlyValIleIyGly...ThiGlyAsnGlyPheGluPro	97
237	TGTAGCTAACGGCATTTTGAAGGGTGTAGTCCAACTCAATTTCCACCA	286
97	snGlyLysIleAspArgValIserMetAlaSerLeuLeuValGluAlaIyr	113
287	ACAATCTTAACTCGTCTGTGAAGCGCTAAAAATTTTGTAGGAAGCATTC	336
114	LysLeuAspThrLysValAsnGlyThrProAlaThrLysPheIyAspLe	130
337	GGTTTAAACAGCTTAACCAAT.....CTTACCGAATTTGGTCGACGC	377
130	uGluThrLeuAsn...TrrGlyLysGluLysAlaAsnIleuValGluI	146
378	TTCTCAAGTAAACCTTGGGCTAAAAAATTACTACAANTGCCAGTACTGA	422
146	eugIyIleSerValGlyThr...GlyAspGlnTrrGluProLysLysThr	161
128	ACGGCAATTTGCAAGGTACTGTATGCAACCAACTTAACCTTACACACTCA	477

```

162 valThrLysAlaGluAlaIleGlnPheIleAlaLysThr...AspLysGln 177
    ::::: ::::: ::::: ::::: :::::
478 ATCACTCGCAAGACTTTCACAGTGTTCAAACGTAAGTGCACAAAGT 527
177 nPheGlyThrGluAlaIleLysValGlnSerAlaLysAlaValThrThc 194
    ::::: ::::: ::::: ::::: :::::
528 TGAAGGTCAAACTCCGAAGAAGACCATTTGTTAAAGCTATCAACACA 577
194 LysValGlnValLysPheSerLysAlaValGlnLysLeuThrLys... 209
    ::::: ::::: ::::: ::::: :::::
578 CAACGTGTGAAGTACATTCGAAGAAGAGTACTACCTTCACAGACT 627
210 .....GluAspIleLysValThrAsn.....LysAlaAs 219
    ::::: ::::: ::::: ::::: :::::
628 AACTTCMAAATCGAAGTTTACAAATTAATAATGCTTGTAAACAAAC 677
219 nasAspLysValLeuVal.....LysG 227
    ::::: ::::: ::::: ::::: :::::
678 AAACAAAAGATTGTGTATTAACTAGTCAACCTCAACAGCTGATTAAG 727
227 LysValThrLeuSerGluAspLysArgSerAlaThrValGlnLeuThrSer 243
    ::::: ::::: ::::: ::::: :::::
728 AGTATGTTTAACTCTTGAC.....GGCGAAACATCGGTGGCTTAA 771
244 AsnLeuAlaLysGlnThrThrValAspValAsnLysValGlnLys 260
    ::::: ::::: ::::: ::::: :::::
772 GGTGGTGGTGT.....GTAGTTCACAACTAAAGTTCGAAC 806
260 sThrGluValAlaVal...GlySerLeuGluAlaLysThrIleGlnMet 276
    ::::: ::::: ::::: ::::: :::::
807 ACTATCTTCTGCAGTTCAGAGTAACTGTGTCAAGAGTAAAGTTCAAG 856
276 LAspGlnThrValValAlaAspGluProThrAla.....LeuGln 289
    ::::: ::::: ::::: ::::: :::::
857 CTAAAGTACTGTGCTGGAAGCTCAATCAAGCTGATTCCTCTACT 906
290 PheThrValLysAspGlnAsnGlyThrGluValValSer..... 302
    ::::: ::::: ::::: ::::: :::::
907 TTCACGTACCAAGTAAACAATGATGGCTGTACCAACATTTAACAG 956
302 ..... 302
957 TGAAGCTTTAACAAAGCAAGAGGTATGCCAACATACCTTACACTCGTT 1006
303 ..ProGluGlyIleGluPheValThrProAlaAla..... 313
    ::::: ::::: ::::: ::::: :::::
1007 ATTAAGAAGTACTGATCAAGTAACTGTTATGCAACACTGTGATCGTCT 1056
314 .....GluLysIleAsnAlaLys 319
1057 AATTCTCACTGGTTATGTATTCGGGGGTGATACAAATTCCTTCACT 1106
319 sGlyGluIleThr.....LeuAlaLysGlyThrSerThrThrThv 332
    ::::: ::::: ::::: ::::: :::::
1107 TCAAGACAGTACTACAGGTGCTTCAGTTATATATGTGCAACAAACTT 1156
332 sLysAlaValThrLys.....LysAspGlyLysVal..... 342
    ::::: ::::: ::::: ::::: :::::
1157 ACAAAAGTACTATATAAAACCTAAACCTGTAACCAAGCAACAAACAA 1206
343 .....ValAlaGlnSerLysGln 348
1207 ACATTTAATGTTGTTTGTAGAAAACATGANTGTACTTCTGATAAAGT 1256
348 uValLysValSerAlaGlnGlyValAlaValAlaSerIleSerAsnTrpT 365
    ::::: ::::: ::::: ::::: :::::
1257 ACCAAATGCTACAGTTAATGGCTAAAGCATTTACAATTAAGCAATGGTA 1306
365 hValAlaGlnGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
    ::::: ::::: ::::: ::::: :::::
1307 CAGCTTTAGAC.....GCTGCTCAAAATTACAAACAGATTCTTAAGTGAA 1350
382 AsnAsnLysValThrGlnGlyAspAsnAlaThrValGlnValGlnLeuLys 398

```

```

1351 GCTACATTCACAGTTTCTGTACTATGACGCTGAACCCAGTAGTATA 1400
    ::::: ::::: ::::: ::::: :::::
398 sAsp.....GlnPheAsnAla 403
    ::::: ::::: ::::: ::::: :::::
1401 TGATCTACACAGCACTAACATATGACTTCAAATATAAATAATATGTGCT 1450
404 .....ValThrThrGlyLysValGlnThrGlnSerLeuAsnThr 416
    ::::: ::::: ::::: ::::: :::::
1451 CTGCTTTACAACTACTGCTTCTTAAGTAACTTGGCTGCTTTCACAGA 1500
417 GluValAlaVal.....ValAspLysAlaThrGlyLysValThr 429
    ::::: ::::: ::::: ::::: :::::
1501 GAGTATACATTTGAGTAACTGCTGCTGATATATGCT...GGAGAAGTGC 1547
429 rValLeuSerAlaGlyLysAla...ProValLysValThrValLysAsp 445
    ::::: ::::: ::::: ::::: :::::
1548 TGCATTTGGCTGCTACTAACGCTGCGGAATACAAAGTTATTGTAAAGATA 1597
445 eLysGlyLysAlaLeuValSerHisThrValGlnIleGlu..... 458
    ::::: ::::: ::::: ::::: :::::
1598 AAGCTGTAACTTACTTAATAAATGAATCGTTAATGTTCATTAATGA 1647
458 ..... 458
1648 GATAAAGATCGTATATTTCACAGTTACAAATGCTAAATTCGTGATAC 1697
459 .....AlaPheAlaGlnLysAlaMetLysAspIleL 469
    ::::: ::::: ::::: ::::: :::::
1698 TGATCCAGTACTGCAATATCTTCAACAGCGATTAACCAAAACAATCT 1747
469 yLeuGlnLysThrAsn..... 474
    ::::: ::::: ::::: ::::: :::::
1748 CTGTA...AAACAATGATAAAGTGAAGCTACATTTGTATCGTCT 1794
475 .....ValAlaLeuSerThrLysAspValThrAs 484
    ::::: ::::: ::::: ::::: :::::
1795 GATACAGTAAACGATTATGCAACACCAATGCTGATGATTAATAATAC 1844
484 PheLysValLysAlaProValLeuAspGlnThrGlyLysGlnPheThr 501
    ::::: ::::: ::::: ::::: :::::
1845 TTCTGATGCAAAACAAGCGACCTGATGAAGTCAACCAAAAGCATGG 1894
501 LAspValThr...ValLysValLeuAspLysAspGlyLysGlnLeuLys 516
    ::::: ::::: ::::: ::::: :::::
1895 CACCAATCTCTTACTTCCAGACCATATCTGATGCTGACGTACAA 1944
517 GluGlnLysLeuGluAlaLysThrValAsnArgGlnLeuValLeuAsnAl 533
    ::::: ::::: ::::: ::::: :::::
1945 GCATACAAA.....AAATCAGATCTTATATAAGCTGTAACTAAGTT 1985
533 sAlaGlyGlnGluAlaGlyAsnThrThrValValLeuThrAlaLysSerG 550
    ::::: ::::: ::::: ::::: :::::
1986 TGATGTTCTGAACCTGCAGTATTTTCAGACGAATTAAGTAACCAAGCG 2035
550 LysGlnLysGluAlaLysAlaThrLeuAlaLeuGlnLeuLysAlaProGly 566
    ::::: ::::: ::::: ::::: :::::
2036 GCAAAAAGTAACTGCTACTCTTATT.....AAGAAGCA 2070
567 AlaPheSerLysPheGlnValArgGlyLeuAspThrGlnLeuAspLysTy 583
    ::::: ::::: ::::: ::::: :::::
2071 ACTTATACAAATCTCAAAATCGGTGCTAATGATATTAAGTAGATTAACA 2120
583 rValThrGlnGluAsnGlnLysAsnAlaMetThrValSerValLeuProV 600
    ::::: ::::: ::::: ::::: :::::
2121 AGTTATCTCACCAAAATGCTAGCTACACAGTACTATATAGCT...ACTT 2167
600 sAlaPheAlaAsnGlyLeuValLeuLysGlyAlaGlnAlaAlaGln..... 614
    ::::: ::::: ::::: ::::: :::::
2168 TATCTTCAACAGAACTGTATTAACCTGCTAAGAAATTAAGAGTTACT 2217
615 .....LeuLysValThrThrThr..... 620
    ::::: ::::: ::::: ::::: :::::

```

```
2218 TCAGTGCATGTTAAAAACAATGCTGTTAAAGTAAATTGCTACAGGTATTGC 2267
621 ...AsnLysGluGlyLysGluValAlaSerPalaThrAspAlaGlnValThrV 636
    ||| :|||:|||||
2268 TGTAAATACACAGCGTAAAGACATGATTTACTGCTAAAGACGTACAG 2317
636 a1.....GlnAsnSerValIleTh 643
2318 CTACATTCACAGCTACAATAATGATTCACAACTCTTACACTGCTAGCT 2367
637 .....GlnAsnSerValIleTh 643
2368 ACTCAATTCATACAGCTGATTCGCTTCAACACAGCACTCTATTGGTT 2417
643 rValGlyGln.....GlyAlaLysAlaGlyLysThrTYL 655
    :|||:|||||
2418 TGCCTGTAATAAACCCAGTGAATATGCTGCTATACAGCAACAATATA 2467
655 ys..... 655
2468 AATACCTTCGAGCTAATGTAATGTAATGTAATGTAATGTAATGTAATG 2517
656 .....ValThrVa 658
2518 GCATTATTACTCAATATGCAACTGAGGCCAATAAGTACATCTCATTA 2567
658 lValLeuAspGlyLysLeuIleThrHisSerPheLysValValAspT 675
    :|||:|||||
2568 TAAAGTACATGCTGATACAGTTAC.....TTTAAAGTAATTAAT 2607
675 hrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeu 691
    |||:|||||
2608 .....GCTGTAAATTCCTTCAACTGAGCACTGATC 2634
692 LysGluValAlaProAsn.....AlaAspLeuLysAlaIleLeu 705
    ||| :|||:|||||
2635 AAACAGCTGCTCCACACACAGCACTGCTCACTGCTGCGCATTAAC 2684
705 uAsnIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThr 722
    :|||:|||||
2685 ATTAACACACAGCTGCTGCTGTTAGTTGATTAACACGCACTGCACTA 2734
722 lAserAsnValGluPheValSerAlaAspThrAsnValValAla..... 736
    :|||:|||||
2735 CTTAGGAATTCATTAAGTACGATCTTAATGTAAGTCAACAACACT 2784
737 ...GluAsnGlyThrValGlyAlaLysGlyAlaThrSerIleTyValLy 752
    :|||:|||||
2785 GTTGATCTGCACTGTTTCATTAAGAATAGTCAATTAATTCATTA.. 2832
752 sAsnLeuThrValValLysAspGlyLysGluGlnLysValGluPheAspL 769
    :|||:|||||
2833 .TCCTTACATTAAGTGAACAGTGTAAATACAGGTGA...TTTGCTA 2878
769 yAlaValAlaGln.....ValAla 774
    :|||:|||||
2879 CAACGTTCACAGCTGCTGATTAATCTTTAAAGTCACTGATACATTAACA 2928
775 ValSerIleLysGluAlaLysProAla 783
    :|||:|||||
2929 GTTACTTATGCAGATGCTAAAAATGCT 2955
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-682-517-7
seq_documentation_block:
; Sequence 7, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-682-517-7

alignment_scores:
  Quality: 378.50      Length: 1009
  Ratio: 0.762        Gaps: 44
  Percent Similarity: 49.257  Percent Identity: 22.101

alignment_block:
US-09-754-947-1 x US-08-682-517-7 ..
Align seg 1/1 to: US-08-682-517-7 from: 1 to: 4197

16 AspSerIleAsnTyLLeuValGluLysGluValAlaValLysGlyAsnAspLy 32
   :|||:|||||
227 GAAGCAGTTCAGCTTTAGTTGACCAAGCGCTAATCCAGGTGATACAA 276
32 sGlyMetPheGluProGlyLysGluLeuThrArgAlaGluAlaThrW 49
   :|||:|||||
277 CGGCAATCTCAACCCACTTAACACAGTCACTGTCACACAGCTGACAAA 326
49 etMeAlaGlnIleLeuAsnLeuProIleAspLysAspAlaLysProSer 65
   :|||:|||||
327 TCTTCACAAAAGCTTTAGAAATTAAGCTAACGGAGATGA.....AAC 370
66 PheAlaAsp...SerGlnGlyInTrpTyThrProPheIleAlaAlaVal 81
   ||| :|||:|||||
371 TTCAAGACGCTGAAGACGCGCTGTGATACCAACTCAATCGCTGCTGT 420
81 lGluLysAlaGlyValIleLysGly...ThrGlyAsnGlyPheGluPro 97
   :|||:|||||
421 TGTAGCTAACGCGCATTTTGAAGGTGTAGTCAACTGAATTTGCACCAA 470
97 snGlyLysIleAspArgValSerMetAlaSerLeuLeuValGluAlaTy 113
   :|||:|||||
471 ACAAATCTTTAAGCTGTTCTGAAGCTGCTAAATTTTAGTAGAAGCATTC 520
114 LysLeuAspThrLysValAsnGlyThrProAlaThrThrLysPheLysAsp 130
   :|||:|||||
521 GGTTAGAAGGTGAAGCAGAT.....CTTAGCGAATTTGCTGACAGC 561
130 uGluThrLeuAsn...TrpGlyLysGluLysAlaAsnIleLeuValGlu 146
   :|||:|||||
562 TTCTCAAGTAAACCTTGCGCTAAATAAATCTTAGAAATCCACAGTGA 611
146 euGlyIleSerValGlyThr...GlyAspGlnTrpGluProLysLysThr 161
   :|||:|||||
612 ACGGCAATTTGGAAGTACTGATGCAACAACCTTAACCTTAACACATCA 661
162 ValThrLysAlaGluAlaAlaGlnPheIleAlaLysThr...AspLysG 177
   :|||:|||||
662 ATCACTGCTCAAGACTTGTGACACTAGTGTCAACAGTACAGTTGACAAAG 711
177 nPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrTh 194
   :|||:|||||
712 TGAAGGTGAACCTCCAGAAAGACGACATTTGTTAAAGCTATCAACAAC 761
194 lntLysValGluValLysPheSerLysAlaValGluLysLeuThrLys... 209
   :|||:|||||
```



```
637 .....GlnAsnAsnSerValIleth 643
2552 ACATCAATTCATACAGCTGATTCGTGTTCAACAGCAACTCTATTGGTT 2601
643 rValIGLYGln.....GlyAlaLysAlaGlyGlnThrTyrL 655
2602 TCGTGGTAACCAACAGTGAATATGCTGGTGTATCATGACGCAAAACATATA 2651
655 ys..... 655
2652 AATACCTCGAGAGCTAATGTAATGAGTAATTTGGTGAACGGCATGGAA 2701
656 .....ValThrVa 658
2702 GCATTATTAACTCATATGCAATGCAAGCCAAAGCAATCAATCATATA 2751
658 lValLeuAspGlyLysLeuIleThrThrHisSerPheLysValAlaLysP 675
2752 TAATGTAGATGGTATACAGTTACA.....TTTAAAGTAATTAGT. 2791
675 hrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrLeu 691
2792 .....GCTGTATATCTTCACACTGAGCATATC 2818
692 LysGluValAlaProAsn.....AlaAspLeuLysAlaAlaLeu 705
2819 AACCAGCTTCTCCACACACAGCAGCTCCACTGCGCATATAC 2868
705 uAsnIleLeuSerValAspGlyValAlaProAlaThrThrAlaLysAlaThra 722
2869 ATTAACCCAGCAGCTGTGTTAGTTAGTTATTAACAACCTCACTAC 2918
722 lAserAsnValGluPheValSerAlaAspThrAsnValAlaAla..... 736
2919 CTGTAGGAATTTCTCATTTAGCTGATGATCTTAATGTAAGCAACAAC 2968
737 ..GluAsnGlyThrValGlyAlaLysGlyAlaLeuThrSerIleTyrValL 752
2969 GTTGATATCTCACTGTTCTTCAATTAAGATGCAATTAATTCATTA.. 3016
752 sAsnLeuThrValValLysAspGlyLysGluGlnLysValGluPheAspL 769
3017 .TCTCTACATTAAGTTGAACGTGCTAATACAGGTGA...TTTGCTA 3062
769 ySalAlaValGln.....ValAla 774
3063 CACTGTGTCAAGCTGTACATTATCTTTAACTGCTGTACATTACA 3112
775 ValSerIleLysGluAlaLysProAla 783
3113 GTTACTTATGCAGATGCTAAATAATGCT 3139

seq_name: /c9n2_6/plodata/2/lna/5B.COMB.seq:US-08-682-517-8
seq_documentation block:
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..184
; US-08-682-517-8
```

```
alignment_scores:
  Quality: 378.50      Length: 1009
  Ratio: 0.762        Gaps: 44
  Percent Similarity: 49.257  Percent Identity: 22.101
```

alignment_block:

US-09-754-947-1 x US-08-682-517-8 ..

Align seq 1/1 to: US-08-682-517-8 from: 1 to: 4197

```
16 AspSerIleAsnTyrLeuValGluLysGlyAlaValLysGlyAsnAspL 32
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
227 GAACGACGTTCAAGCTTTAGTTGACCAAGCGCTAATCCAGCTGCTACTAC 276
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
32 sGlyMePheGluProGlyLysGluLeuThrArgAlaGluAlaLeuThrM 49
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
277 CGGGAACCTTCAACCCACTTAACACAGTAACACTGCTGCACAAAGCTGACAAA 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
49 etheAlaGlnIleLeuAsnLeuProIleAspLysAspAlaLysProser 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 TCTTCACAAAAGCTTTGAATTTAGAACCTTAACGATGATGTA.....AAC 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 PheAlaAsp...SerGlnGlyIntRPTyrThrProPheIleAlaAlaVa 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
371 TTCAAAGACGTGAAGCTGCGCGCTGTGTTACTACATCAATCGTCTGCTGT 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
81 lGluLysAlaGlyValIleLysGly...ThrGlyAsnGlyPheGluProA 97
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 TGTAAGCTTAACGCGCATTTTGAAGCTGTAGTGCACAACTGAATTGCAACCA 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
97 snGlyLysIleAspArgValSerMetLaserLeuValGluAlaTyr 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
471 ACAAAATCTTAACCTGTTCTGAAGCTGCTAAAAATTTAGTAGAAGCAATTC 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
114 LysLeuAspThrLysValAsnGlyThrProAlaThrLysPheLysAspLe 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
521 GGTTTAAGAGTGAGCAGAT.....CTTAGCCAAATTTGCTGTGACCC 561
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 uGluThrLeuAsn...TyrGlyLysGluLysAlaAsnIleLeuValGluL 146
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
562 TTCACAGTAACCACTTGGGCTAAATAATCTTAGAATGCGACTACTCA 611
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
146 euGlyIleSerValGlyThr...GlyAspGlnIlePheLysLysThr 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
612 ACCGCAATTTTCGAAGTACGATGCAACAAACCTTAACCTTACCAACTCA 661
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
162 ValThrLysAlaGluAlaAlaGlnPheIleAlaLysThr...AspLysG 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
662 ATCACTGCTCAAGACTTTCAGCTAGTGTTCAAAAGTACAGTTGCAAGT 711
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
177 nPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrg 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
712 TGAAGGTGAACACGCCAAGACAGACAGCATTTGTTAAAGCTATACACACA 761
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

194 LuLysValGluValLysPheSerLysAlaValGluLysLeuThrLys... 209
    ||||| ||||| ||||| |||||
762 CAACCTGTTGAAGTACATTCGAAAGAGATGTTACTTAACGTTCAAGACGTT 811
210 .....GluAspLeuLysValThrAsn.....LysAlaAsn 219
812 AACCTCAAAATGCAAGGTTTAGAAATTAATAAGCTTCTGTTAAACAAAC 861
219 AsnAspLysValLeuVal.....LysG 227
    ||||| ||||| ||||| |||||
862 AAACAAAAAAGTTGTTGTTACTTAACCTCAAGCTCAAAACGCTGATTAAG 911
227 LuValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSer 243
    ||||| ||||| ||||| |||||
912 AGTATGTTTAACTCTTGAC.....GGCGAAACAATCGTGCGCTTAA 955
244 AsnLeuAlaAlaLysGlnThrTyrThrValAspValAsnLysValGlyLys 260
    ||||| ||||| ||||| |||||
956 GGGTGGCTGCT.....GTAGTTCCAACTAAAGTTGAACT 990
260 sThrGluValAlaVal...GlySerLeuGluAlaLysThrIleGluMet 276
    ||||| ||||| ||||| |||||
991 AGTATCTTCTGCAGTTCAGGTAACCTTGTCAGAGAGTAAGTTCAAG 1040
276 LAspGlnThrValValAlaAspGluProThrAla.....LeuGln 289
    ||||| ||||| ||||| |||||
1041 CTAAAGTAACTGTGCTGAAGGTCATTAAGCTGATTCCTGTTACT 1090
290 PheThrValLysAspGluAsnGlyThrGluValValSer..... 302
    ||||| ||||| ||||| |||||
1091 TTCACCTGACCGATACACATGATGCGTTGTACCAACATTAACAG 1140
302 ..... 302
1141 TCAGCTTTAACAAAGAGAGGATATGCAACATCTTACTACTCGTT 1190
303 ..ProGluGlyIleGluPheValThrProAlaAla..... 313
    ||||| ||||| ||||| |||||
1191 ATAAAGAGGTAAGTATGTAAGTAACTGCTATGCAACTGCTGATCTCT 1240
314 .....GluLysIleAsnAlaLys 319
1241 AATCTCTCACTGGTATGATTCGCGGTGATACATCTTCTGACT 1290
319 sGlyGluIleThr.....LeuAlaLysGlyThrSerThrThrV 332
    ||||| ||||| ||||| |||||
1291 TGAAGAACTAATCAAGGCTGCTTCAATTAATGTCGCAAAACAACCTT 1340
332 AlLysAlaValTyrLys.....LysAspGlyLysVal..... 342
    ||||| ||||| ||||| |||||
1341 ACAAAAGTTACTTAATAAAACCCTAAACTGTTAAACAGAAACAAACAA 1390
343 .....ValAlaGluSerLysG 348
    ||||| ||||| ||||| |||||
1391 ACATTTAATGTTGTTTGTAGAAACATGATGTTACTTCTGATAAGT 1440
348 uValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTPT 365
    ||||| ||||| ||||| |||||
1441 AGCAAAATCTCAAGTTAATGGCGTAAAGCAATTAAGCAATGCTA 1490
365 hrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
    ||||| ||||| ||||| |||||
1491 CAGCTTTAGAC.....GCTGCTCAAAATTAACAGATTTAAAGGCTAA 1544
382 AsnAsnLysValTyrGluGlyAspAsnAlaTyrValGlnValGluLeuLys 398
    ||||| ||||| ||||| |||||
1535 GCTACATTCACAGATTCTGTTACTAATGACCTGTAAGCCGACAGTATA 1584
398 sAsp.....GlnPheAsnAla 403
    ||||| ||||| ||||| |||||
1585 TGATCTACACAGACATAAATAGTACTTCAAAATTAATAATATAGTCAT 1634
404 .....ValThrThrGlyLysValGluTyrGluSerLeuAsnThr 416
    ||||| ||||| ||||| |||||
1635 CAGCTTACAACTACTGTTCTTAAAGTAACTTTCGCTGCTTAAACA 1684
417 GluValAlaVal.....ValAspLysAlaThrGlyLysValThr 429
    ||||| ||||| ||||| |||||
1685 GAGTATACAAATGAGTAACTGCTGCTGATATGCT...GGAGAGTTCG 1731
429 rValLeuSerAlaGlyLysAla...ProValLysValThrValLysAsp 445
    ||||| ||||| ||||| |||||
1732 TGCAAATGGCGTACTAAGCGCGCAATACAAAGTATTTGTAAGAAAGATA 1781
445 eLysGlyLysAlaLeuValSerHisThrValGluIleGlu..... 458
    ||||| ||||| ||||| |||||
1782 AAGCTGTAACTTAAATGAATCGTTAATGTTGCAATTCATGAA 1831
458 ..... 458
1832 GATAAAGATCGTAAATTCAACAGTACAAATGCTAAATTCGTTGATAC 1881
459 .....AlaPheAlaGlnLysAlaMetLysAspIleL 469
    ||||| ||||| ||||| |||||
1882 TGATCCAGATCTGCGATATCTTCACAGCGCATTAAGCAAAACAATCT 1931
469 yLeuGluLysThrAsn..... 474
    ||||| ||||| ||||| |||||
1932 CTGTA...AAAACAATGATAAGGTGAAGCTACATTTGTTATGGTTCT 1978
475 .....ValAlaLeuSerThrLysAspValThrAs 484
    ||||| ||||| ||||| |||||
1979 GATACAGTAAACGATTAATGCAACACCAATCTGATGATTAATTAATAC 2028
484 PLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGluPheThr 501
    ||||| ||||| ||||| |||||
2029 TTCGATCAAAACAGAGCGACCTTGATGAAGGTGAACCAAAACAAGTGG 2078
501 lAProValThr...ValLysValLeuAspLysAspGlyLysGluLeuLys 516
    ||||| ||||| ||||| |||||
2079 CACCAATCTTACTTCCAGCACCATATCTTGATGGCTGACGATACAA 2128
517 GluGlnLysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAl 533
    ||||| ||||| ||||| |||||
2129 GCATACAAA.....AATCAGATCTTAATTAAGCTGTAACGTT 2169
533 aAlaGlyGlnGluAlaGlyAsnTyrThrValLeuThrAlaLysSerG 550
    ||||| ||||| ||||| |||||
2170 TGATGTTCTGAACCTGACAGATTTGCGACGAAATTAGTAACCAAGCG 2219
550 LysGluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGly 566
    ||||| ||||| ||||| |||||
2220 GCAAAAAGTAACTGTAATTAATTAACCTGCTAAGATTAGAAAGTACT 2254
567 AlAphSerLysPheGluValArgGlyLeuAspThrGluLeuLysLys 583
    ||||| ||||| ||||| |||||
2255 ACTTATCAATCTCAATTAAGTGTGCTAATGATTAAGAGTAAACCA 2304
583 rValThrGlnGluAsnGlnLysAsnAlaMetThrValSerValLeuProV 600
    ||||| ||||| ||||| |||||
2305 AGTATCTCAACCAATGCTAGCTACACAGTAACCTAAGAAAGCT...ACT 2351
600 AlAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu..... 614
    ||||| ||||| ||||| |||||
2352 TATCTTACAGAACTGTAATTAATTAACCTGCTAAGATTAGAAAGTACT 2401
615 .....LeuLysValThrThrThr..... 620
    ||||| ||||| ||||| |||||
2402 TCAGTGAATGTTAAACAACTGCTGTTAAAGTAAATTTGCTACAGGATTCG 2451
621 ....AsnLysGluGlyLysGluValAspAlaThrAspAlaGlnValThrV 636
    ||||| ||||| ||||| |||||
2452 TGTTAATACAGACGGTAAAGACATATGATTTACTGCTAAAGAAAGCTAACG 2501
636 Al..... 636
    ||||| ||||| ||||| |||||

```

```

2502 CTACATTCACAGCTACAAATGAAGTCTTACACTGTGTACT 2551
637 .....GlnAsnAsnSerValIleTh 643
2552 ACTCAATTCATACAGCTGATTCTGTTCACAAACAGCAACTATATTGGTT 2601
643 rValGlyGln.....GlyAlaLysAlaGlyIleuThrTYL 655
2602 TGCTGGTAAACCCAGTGAATAATGCTGTATCAGCCAAACATATTA 2651
655 YS..... 655
2652 AATACCTCGAGCTAATGTTATGATGAGATTGTTGGAAGCGCATGGCAA 2701
656 .....ValThrVa 658
2702 GCATTATTAACTCAATATGCAATGAAAGCCAAAGTAACAACTCTCTA 2751
658 lValLeuAspGlyLysIleuIleThrHisSerPheLysValAlaAspT 675
2752 TAAATGATAGTGTATACAGTTACA.....TTTAAAGTAAATTACT. 2791
675 hValProThrAlaLysGlyLeuAlaGluPheThrSerThrSerLeu 691
2792 .....GCTGTTAATTCTTCACTGAAGCTATTC 2818
692 LysGluValAlaProAsn.....AlaAspLeuLysAlaAlaLeuLe 705
2819 AAACCAAGTTGCTCAACACAGCACTCCAACTACTGCGCATTTAAC 2868
705 uAsnIleuSerValAspGlyValProAlaThrThrAlaLysAlaThr 722
2869 ATTAAACACCAAGCTGTGTTAGTTGATTAAACAACCTCAACTAACA 2918
722 lAsnAsnValGluPheValSerAlaAspThrAsnValAla..... 736
2919 CTTTAGCAATTCATAGCTAGTCAGATCTTAATGTAAGGCCAACACT 2968
737 ..GluAsnGlyThrValGlyAlaLysGlyAlaThrSerIleTyVally 752
2969 GTTGATACCTGCAACTGTTTCATTAAGATAGTGCAAAATTAATTCATTA. 3016
752 sAsnLeuThrValValLysAspGlyLysGluGlnLysValAluPheAspL 769
3017 .TCTCTACATTAAGTTGAACCTGGTGAATATACAGGTGTA...TTTGCTA 3062
769 yAlaValGln.....ValAla 774
3063 CAACCTGTTCAAGCTGTACATTAATCTTAACTGCTGTACATTACA 3112
775 ValSerIleLysGluAlaLysProAla 783
3113 GTTACTTATGCAAGTCTTAAAAAATGCT 3139

seq_name: /cgn2.6/prodata/2/1na/5A_COMB.seq:US-08-119-125A-2

seq_documentation_block:
; Sequence 2, Application US/08119125A
; Patent No. 5610011
;
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; APPLICANT: VECHT, Uri
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly
; TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the diag
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centraal Diegeneskundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Lelystad
; STATE:
; COUNTRY: The Netherlands

```

```

; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS v.6.0
; SOFTWARE: Wordperfect v. 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,125A
; FILING DATE: 20-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00054
; FILING DATE: 19-MAR-1992
; APPLICATION NUMBER: NL 9100510
; FILING DATE: 21-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Handal, Anthony H.
; REGISTRATION NUMBER: 26275
; REFERENCE/DOCKET NUMBER: SMITHHE119125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 838-8589
; TELEFAX: (203) 838-8794
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6744 base pairs
; TYPE: Nucleic acid with corresponding amino acids
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus suis type II (pathogenic)
;
; FEATURE:
; OTHER INFORMATION: Extracellular factor related protein (EF*) gene
;
; FEATURE:
; NAME/KEY: promoter -35 region
; LOCATION: bp 66 to 71
;
; FEATURE:
; NAME/KEY: promoter -10 region
; LOCATION: bp 89 to 94
;
; FEATURE:
; NAME/KEY: promoter -35 region
; LOCATION: bp 153 to 158
;
; FEATURE:
; NAME/KEY: promoter -10 region
; LOCATION: bp 176 to 181
;
; FEATURE:
; NAME/KEY: ribosome binding site
; LOCATION: bp 350 to 356
;
; FEATURE:
; NAME/KEY: signal peptide
; LOCATION: bp 361 to 498
;
; FEATURE:
; NAME/KEY: start of repetitive units R1-R11
; LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,
; LOCATION: 5065, 5293, 5521:
;
; FEATURE:
; NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences
; LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,
; LOCATION: 5128, 5356, 5584:
;
; FEATURE:
; NAME/KEY: dyad symmetry regions
; LOCATION: from bp 6554 to 6566 and from bp 6571 to 6583
;
; FEATURE:
; NAME/KEY: dyad symmetry regions
; LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644
;
; US-08-119-125A-2

```

```

alignment_scores:
Quality: 245.50      Length: 947
Ratio: 0.557        Gaps: 41
Percent Similarity: 46.568      Percent Identity: 21.225

```



```

4799  || ||| ..... ||| ..... |||
      GTGTTGCGCCATTGCGAAGATGTT...CTTGACGCGAGCTAAACAGAT 4845
464  AlamelAspIleLysLeuGluThrAsnValAlaLeuSerThrLy 480
      ||| ..... ||| ..... ||| ..... |||
      GCTAAGATTAAGATGCTTAAGAA...TCCGACGCTGCTAAGTCCGCAAT 4892
480  sAspValThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyL 497
      ||| ..... ||| ..... ||| ..... |||
      TGACCGCGAAT.....CCAAACTTGACACATCGACAGAGA 4924
497  ysgIuPheThr.....AlaProValThrValLys 506
      ||| ..... ||| ..... ||| ..... |||
      AGGAATCAGCTAAGAAACAGTGTGCTGATGCTAAGAGCTCGACAGAT 4974
507  ValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGluAlaLy 523
      ..... ||| ..... ||| ..... ||| ..... |||
      GCAATTGATGCTTCAACAAGATCCAGTCCAGGCAATCCGCGACAGACAA 5024
5025 sTyrValAsn.....ArgGluLeuValLeuAsnAlaAlaGlyGlnGluA 538
      ||| ..... ||| ..... ||| ..... ||| ..... |||
      AGCGCTAAGTTCATCGCCCAAGATGTTCTTGACGCGAACAAGATG 5074
538  laGlyAsnTyrThrValValLeuThrAlaLysSerGlyGluLysGluAla 554
      ||| ..... ||| ..... ||| ..... ||| ..... |||
      CTAAAGAAC.....AAGATTGCCAAAGAGTGGC 5103
5075  LysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysph 571
      ||| ..... ||| ..... ||| ..... ||| ..... |||
      GCAGCTTAAGAAAGCAATGATGCCAATCG..... 5133
5104  egluValArgGlyLeuAspThrGluLeuAspLysTyrValThrGluGluA 588
      ..... ||| ..... ||| ..... ||| ..... |||
      AACTTATCAGATCGACAGAAAGACTTCTAAGAAAGCGTAGATGCGA 5182
5134  sngIuLys.....AsnAlaMetThrValSerValLeuProValAsp 601
      ..... ||| ..... ||| ..... ||| ..... |||
      ATGCTTAAAGCTACGACAGATGCAATTGATGCTTCAACAGTCAGTCGAA 5232
5183  AlaAsnGlyLeuValLeuLysGlyAla..... 610
      ||| ..... ||| ..... ||| ..... ||| ..... |||
      CGCGCATCGCGCGAGACAAAGCGTAGTTCATCGCCCAAGATGTTCT 5282
5233  GluAlaAlaGluLeuLysValThrThrAsnLysGluGlyLysGluV 627
      ..... ||| ..... ||| ..... ||| ..... |||
      TGACCGCAGCG.....AAACAAGATGCTAAGATTAAGATTGCTAAGAAGAT 5326
5283  aLAspAlaThrAspAlaGlnValThrValGlnAsnAspSerValIleThr 643
      ||| ..... ||| ..... ||| ..... ||| ..... |||
      CCGAGCGTGTGTAAGTCAAGCCATTGACGCGAATCCAACTTGACAGATGCA 5376
5327  ValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThrValValLe 660
      ..... ||| ..... ||| ..... ||| ..... |||
      GAGAGAGATCGACTAG..... 5394
5377  uAspGlyLysLeuIleThrThrHisSerPheLysValValAspThrAlaP 677
      ..... ||| ..... ||| ..... ||| ..... |||
      .....AAGCGGTAGAT..... 5406
5395  roThrAlaLysGlyLeuAlaValaGluPheThrSerThrSerLeuLysGlu 693
      ..... ||| ..... ||| ..... ||| ..... |||
      ..... 5406
5406  ValAlaProAsnAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerVal 710
      ..... ||| ..... ||| ..... ||| ..... |||
      .....GCAAGTCTAAGCTGCG.....AC 5426
5407  lAspGlyValProAlaThrThrAlaLysAlaThrAlaSerAsnValGlu. 726
      ..... ||| ..... ||| ..... ||| ..... |||
      AGATGCAATTGATGCTTCAACAAGTCCAGTCGAAAGCGCAATCGCGAGAG 5476
5427  pheValSerAlaAsp..... 731
      ..... ||| ..... ||| ..... ||| ..... |||

```

```

5477  ACAAAGCGGTAGCCCGCATCGCCCAAGACATTCTTGATGCCCGCAACAA 5526
732  .....ThrAsnValValaLagLusnnglyThrValGlyAlaLysGlyAl 746
      ||| ..... ||| ..... ||| ..... |||
      GATGCTAAGAACAAAGATGCTTAAGAG.....GCAGAAATCCGC 5564
5527  athrSerIleTyrValLysAsnLeuThrValValLysAspGlyLysGlu. 762
      ..... ||| ..... ||| ..... ||| ..... |||
      TAAGTTCAGTCATGACTCCAAATCCGAACTTGACAGATGCGAGCTAAGGAG 5614
5565  GluLysValaGluPheAspLysAlaValaGluValaLaval 775
      ||| ..... ||| ..... ||| ..... ||| ..... |||
      CCGCTTAAATCTGAATGTGATTAAGCTGTGAGAGAACCGATT 5655
5615  seq_name: /cgn2/6/plodata/2/lna/6b_cowb.seq:US-08-913-942-3
seq_documentation_block:
: Sequence 3, Application US/08913942
: Patent No. 6200578
: GENERAL INFORMATION:
: APPLICANT: St. Geme, Joseph
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr Hochbach Test Albritton & Herbert LLP
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/913,942
: FILING DATE: 29-DEC-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/409,995
: FILING DATE: 24-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/4031
: FILING DATE: 22-MAR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Vance, Dolly A.
: REGISTRATION NUMBER: 39,054
: REFERENCE/DOCKET NUMBER: A-61053-1/RPT/RMS/DAV
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7291 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: DNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 163..7221
: US-08-913-942-3
alignment_scores:
  quality: 240.50      length: 850
  ratio: 0.584         gaps: 41
Percent Similarity: 48.471  Percent Identity: 22.706
alignment_block:

```

US-09-754-947-1 x US-08-913-942-3 ..

Align seg 1/1 to: US-08-913-942-3 from: 1 to: 7291

```

23 GluysGlyAlaValLysGlyAsnAspLysMetPheGluProGly 39
   |||||
397 GAATAAGATTACAGAAATTCATAATTCGGGAATATTTTCACACATAA 446
   |||||
39 sGluLeuThrArgAlaGluAlaAlaThrMetLeuLaglinLeuAsn 56
   |||||
447 AGGAGTACTAAAAAGCCGAGCAATCACCCTCAAAAGCCGCGAC...AAC 493
   |||||
56 euproLeaspLysaspAla.....LysProserPheAla 67
   |||||
494 TGAATAATCAACAAACACCGATGAAGCAACCAATGCCAGTACTTCACC 543
   |||||
68 AspSerGlyGlyLeuThrPheProPheAlaAla...AlaValGlu 83
   |||||
544 TACTCCCTCAAAAAGACCTCAACGATCTGACCAGTGTGCACTGAAAA 593
   |||||
83 salaglyVal.....IleLysGlyThrGlyVal 92
   |||||
594 ATTATGCTTTGGCGCAACGCGATTAAGTTGATATTACCAAGTATGCA 643
   |||||
92 snGlyPheGlu.....ProAsnGlyLysIle..... 100
   |||||
644 ATGGCTTGAAATTCGCCAAACAGTAAACGGAATGTTTCATTGAAATGT 693
   |||||
101 .....AsparGlySerMetAlaSerLeuLeuVal 111
   |||||
694 TTGCATTCACTTTGCTGATGCGGTACGAAATACAGGTGTGTTAACTTC 743
   |||||
111 uAlaLeuLysLeuAspThrLysValAsnGlyThrProAlaThrLysPhe 128
   |||||
744 ATCAAGTTTACACCTATATGATGTTGAAAAACAGACGCTGCACACTGTA 793
   |||||
128 ysaAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
   |||||
794 AAGATGTTTAAATGCAAGTGTGAAACATTAAGGTCTAAAAACT..... 837
   |||||
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProLysLysTh 161
   |||||
838 .....GCTGGAGGTAAATGTTGAGAGCTGTGATTT 866
   |||||
161 rValThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAspLysGln 178
   |||||
867 AGTGTCCGCTTATATATATGTTGAATTATTATACAGCGCATTAACACGC 916
   |||||
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrGln 194
   |||||
917 TTGATGTTGTTTAAACAGCTAAAGAAAACGGTAAACCA..... 954
   |||||
195 LysValGluValLysPheSerLysAlaValAlaGluLysLeuThrLysGluAs 211
   |||||
955 ...ACCGAAGTGAATTCACACCGAAACCTCTGTATCAAAAGAAAAGA 1001
   |||||
211 pIleLysVal...ThrAsnLysAlaAsnAsp.....LysValLeuVal 225
   |||||
1002 CGGTAAGTATTACTCGAATAAGCAATACACACAATTAAGTTTACAA 1051
   |||||
225 aLysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
   |||||
1052 GTAAACACCGCATCTGATTAACAGATGAGGGTAAATGGCTTAGTCACTGCA 1101
   |||||
242 TyrSerAsnLeuAlaAlaLysGlnThrThrValAspValAspLysVal 258
   |||||
1102 AAAGCTGTGATGTGATCT.....GTCAACACAAGGC 1130
   |||||
258 lGly.....LysThrGluValAla..... 264
   |||||
1131 TGCTTGAGAGTAAACAACTACTGTAATGTCAAAATGGCGACTTCG 1180
   |||||
265 ...ValGlySerLeuGluAlaLysThrIleGluMetAlaAspGlnThr 279

```

```

1181 CAACGTTCGCTCAGCACAATAATGTAACCTTTGAAGATGGCGCATGTACA 1230
   |||||
280 ValValAlaAspGluProThrAlaLeuGlnPheThrValLysAspGluAs 296
   |||||
1231 ACAGCGTCA.....GTAACCTAAAGATCTCTAA 1256
   |||||
296 nGlyThrGluValAlaSerPro.....GluGlyIleG 307
   |||||
1257 CGCAATGGCATCACCACCTTAAGTACGACCGCAAGTGGCGAGCGCTTGA 1306
   |||||
307 LuPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIleThr 323
   |||||
1307 AATTT.....GATACGATAAAAAATC..... 1329
   |||||
324 LeuAlaLysGlyThrSerThrThrValLysAlaValTyrLysLysAspG 340
   |||||
1330 GTTGCAGATACGACCCGACTTACTGTGACAGCT.....G 1364
   |||||
340 yLysValAlaAlaGluSerLysGlu.....ValLysValSerAlaG 354
   |||||
1365 TAAGTAGTGTGAATTCCTAAAGAGATGACAAAGAAAAAATCTGTTAATG 1414
   |||||
354 LuGlyAlaAlaValAlaSerIleSerAsn.....TrpThrValAlaGlu 368
   |||||
1415 CAGCGCATTTGTTAACAAGCTTTAGTAACTTAAGTGTGAAGCAAAAGCT 1464
   |||||
369 GluAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAsnLysVal 385
   |||||
1465 GAGCGTATCTGATGTGTGCGCTTGAAGGGATTTCAAAACCAAGAGAGT 1514
   |||||
385 lTyrGluGlyAspAsnAlaTyrValGluValGluLeuLysAspGlnPhe 402
   |||||
1515 CAAGCGCGCAACCGGTA.....ACCTTTA 1540
   |||||
402 snAlaValThrThrGlyLysValGluTyrGluSerLeuAsnThrGluVal 418
   |||||
1541 AAGCGGCAAGAACTTAAAGTGAACAGAGATGCGCAACTTACTTAT 1590
   |||||
419 AlaValAlaAspLysAlaThrGlyLysValThrVal..... 430
   |||||
1591 TCACGTCAAGATGCTTAAACGCGTTAAACGAGCTTACTTAGTGTGTAC 1640
   |||||
431 LeuSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
   |||||
1641 AACTAAATGGCGGAATGATGCGAAACCGTCAACAAACAGCGTTAA 1690
   |||||
445 .....SerLysGlyLysAlaLeuValSerHisThrValGlu 456
   |||||
1691 CCATCAGCGCCAGAGTAAATGCGGTACGACAGGTACAAACCATTCAGC 1740
   |||||
457 lIleGluAlaPheAlaGlnLysAla...MetLysAspIleLysLeuGlu 472
   |||||
1741 GTAACCAAGATGGCATTAAGACAGGTAAATTAAGCTAAT..... 1779
   |||||
472 sThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLysA 489
   |||||
1780 .ACTAATTTGCGAGTGTGTTAAGAGCTTATGACGAT..... 1815
   |||||
489 laproValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrVal 505
   |||||
1816 .....GCGAATTTT 1824
   |||||
506 LysValLeuAspLysAspGlyLysGluLeuLysGluGln..... 518
   |||||
1825 GATGTTTAAATTAACCTGCACTGATTAATTAACACAGCTGAAGATGC 1874
   |||||
519 .....LysLeuGluAlaLysTyrValAsnArgGlu...L 529
   |||||
1875 TTATAAGGTTTATTAATCTAAATGAAGAAAAAATGCAATTAACAAACGT 1924
   |||||
529 euValLeuAsnAlaAlaGlyGlnGluAlaGlyAsnTyrThrVal..... 543
   |||||

```

```

1925 TGGTACTGACAGCAGCGCGGACTGATGAGCGATTACGTAATTGGCT 1974
544 ..ValLeuThraLysSerGlyLysGluAla.....LysAlaThr 557
      |||.....|||.....|||
1975 TGGGTACTATCAACCAAAACGTAACGAAAGCAATCAAGTTAA 2024
557 rLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPheGluValA 574
      ||| |||.....|||
2025 ACAAGCTGAGCAAGTCTTACCGGGGCC..... 2055
574 rGlyLeuAspThrGluLeuAspLysThrValThrGluGluAsnGluLys 590
      ||| |||.....|||
2056 ..GGTCTGCTACGGTTACTTCCAA.....TCGAAACGGTTAA 2094
591 AsnAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLe 607
      |||.....|||.....|||
2095 CATACGATTACCGTTAGTGTCGGAACCTAAGCGGATTCGGCTCTGA 2144
607 uLysGlyAlaGluAlaGluLeuLysValThrThrAsnLysGluG 624
      |||.....|||.....|||
2145 AAAAGATGGCCATCTATTAGCTCAAGTGAATATCAAAAC..... 2187
624 LysGluValAspAlaThrAspAlaGluValThrValGluAsnAsnSer 640
      |||.....|||
2188 .....ACTGATATGTTTAACTGTTGTAATAT... 2217
641 ValIleThrValGlyGlyAlaLysAlaGlyLysThrLysValThr 657
2217 ..... 2217
657 rValValLeuAspGlyLysLeuIleThrThrHisSerPheLysValValA 674
      |||.....|||
2218 .....GCTACTGCTGCACATAAGCTGCTTAAACCTGTTA 2254
674 sPThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSer 690
      |||.....|||
2255 AAACGTGAGCGCACT.....GATGCAGATCGCGGTAAAGTAACT 2292
691 LeuLysGluValAlaProAsn..AlaAspLeuLysAlaAlaLeuLys 706
      |||.....|||
2293 GTAAAGATGCTACTGCTAATGACGCTGATGAAGAGTGCACACTGTAA 2342
706 nIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrAla 723
      |||.....|||
2343 AGATGTT.....GCACGCAATTAAATAGTCGGCGGA 2374
723 eAsnValGluPheValSerAlaAspThrAsnValAlaIleGluAsnGly 739
      |||.....|||
2375 CTTTGTGCAAAACAGAGATTAACTACCTCTATTGATGAAGATATCTCT 2424
740 ThrValGlyAlaLys.....GlyAlaThrSerIleTy 750
      |||.....|||
2425 ACAGATTAACGCAAAAGATGACGCACTTAAGCGGCGATACCTTAACCT 2474
2475 rVal.....LysAsnLeuThrValValLysAspGlyLysGluGlnLys 765
      |||.....|||
2475 TAAAGCAGGTAAACCTGAAAGTTAAACGTGATGAAAA.....AATA 2518
765 aLGlupLysAlaValGluValAlaValSerIleLysGluLysAlaLys 781
      |||.....|||
2519 TTACTTTTGACTTGGCGAAAAAC.....CTTGAAGTAAAAAAGTCCGAAA 2562

```

seq_name: /c9n2.6/plodata/2/lna/5A_COMB.seq:US-08-409-995-3

seq_documentation_block:

; Sequence 3. Application US/08409995

; Patent No. 5646259

; GENERAL INFORMATION:

; APPLICANT: Barenkamp, Stephen I.

; TITLE OF INVENTION: St. Geme III, Joseph W.

; NUMBER OF SEQUENCES: 6 Haemophilus Adhesion Proteins

; CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; US-08-409-995-3

```

```

alignment_scores:
  Quality: 238.00      Length: 850
  Ratio: 0.580        Gaps: 42
  Percent Similarity: 48.235      Percent Identity: 22.824

```

alignment_block:

US-09-754-947-1 x US-08-409-995-3 ..

Align seg 1/1 to: US-08-409-995-3 from: 1 to: 5738

```

12 HistrpGlyIleAspSerIleAsnTyrlleuValGluLysGlyAlaVally 28
   |||.....|||
256 AATTGGGGAATA.....TATTGACACATAAAGAGTACTGTAA 293
28 s.....GlyAsnAspLysGlyMetPheGluP 37
   |||.....|||
294 AGCCGACCATACACCCCAAGCCGCGACACACTGAAATTAACAAA 343
37 rGlyLysGluLeuThrArgAlaGluAlaIleThr.....MetAla 51
   ||| |||.....|||
344 ANACGATGAAGAAAGNACCAATGACGTAAGTCTACCTGCTGAAAAA 393
52 GlnIleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAs 68
   |||.....|||
394 GACCTCACAGATCTGACCAAGTGTTCGCAACTGAAAAATTATGTGGGCG 443
68 pSerGlnGlyGlnTrpTyThrProPheIleAlaIleValGluLysAlaG 85
   |||.....|||
444 AAACGGCATAAA.....CTTGATTTTACCACTGATGCA 478
85 lValIleLys.....GlyThrGlyAsnGly...PheGluProAsnGly 98
   |||.....|||
479 ATGCTTGAAATTGGCGAAACAGCTAACGAAATGTTCAATTGAAATGCT 528
99 .....LysIleAspArgValSerMetAlaSerLeuValGlu 111
529 TTGATTCACTTGCCTGATGCGGTAAAGATACGATGTTAAGTTTC 578
111 uAlaTyrlLysLeuAspThrLysValAsnGlyThrProAlaThrLysPheL 128
   |||.....|||
579 ATCAAGTTTACACTATGATGTTGAAAAAACAAGACACTGCACACTGTA 628

```

402 snAlaValThrThrGlyLysValGluTyrGlnSerLeuAsnTrpGluVal 418
::: :::: |||||::: :::: :::: ||
1376 AAGCGGCACAAAGATTAAAGTAGCAAGAGGTGGTCGAACCTTTACTTAAT 1425
AAlaValAlaAspLysAlaThrThrGlyLysValThrVal 430
1426 TCACGCAAGATGCTTTAACGGGTTTACAGCAGCATTTCTTAGTGSTAC 1475
:::::::::: ||| ||||| ::::::::::
431 LeuSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
:::::::::: ||| ||| |||||
1476 AACATAATGCGCAAAATGATGCGAAAACCCGTCATCACAAAGACGGTTTAA 1525
445 SerLysGlyLysAlaLeuValSerHisThrValGlu 456
1526 CCATCACCACCCAGCAAGTAAATGCGGTACGACAGCTACAAACCATCAGC 1575
457 IleGluAlaPheAlaGlnLysAla...MetLysAspIleLysLeuGlu 472
:::: :::: ||||| ||| |||
1576 GTAACCAAAGAATGCGATTAAAGCAGGTAAATTAACCTATT..... 1614
472 sThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLysA 489
||||||| :::: :::: |||
1615 .ACTAATGTTGGCATGGGTTAGAGCTTAAAGCAT..... 1650
489 laProValLeuAspGlnTyrGlyLysGluPheTrpAlaProValThrVal 505
1651 GCCAATTT 1659
506 LysValLeuAspLysAspGlyLysGlnLeuLysGln..... 518
||||| :::: :::: |||||
1660 GATGTTTTAAATAACTCTGCACAACCTGATTTAAATATAGACACGTTGAAGATGC 1709
519 LysLeuGluAlaLysTyrValAsnArgGlu...L 529
1710 TTATAAAGCTTTATTAAATCTAATAATGAAAAAATGCCAATTAACACCGCT 1759
529 euValLeuAsnAlaAlaGlyGlnGluAlaGlyAsnTyrThrVal..... 543
||||| :::: :::: |||||
1760 TGGTAGCTCAGACAGCAGGGCGAGCTGTAGCGATTTACGTAATTTGGGT 1809
544 ...ValLeuThrAlaLysSerGlyLysGlnGluAla....LysAlaThr 557
||||| :::: :::: |||||
1810 TGCGTAGTATCACCAAAACGGTACGAAGAAGAAACCAATCAAGTTAA 1859
557 rLeuAlaLeuGlnLeuLysAlaProGlyAlaPheSerLysPheGluValA 574
||| |||||
1860 ACAAGGTGATGAAGTCCCTTTACGGAGCC..... 1890
574 rgGlyLeuAspThrGlnLeuAspLysTyrValThrGlnGluAsnGlnLys 590
||| ||| :::: |||
1891 ..GGTGCTCTCGAGGTACTTCCA.....TCGAAAACGGTAA 1929
591 AsnAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValle 607
:::::::::: |||||
1930 CATACATTACCGTTAGTGTGGCTCAAACTAAACGGAATTCGGCTGTGA 1979
607 uLysGlyAlaGluAlaAlaGlnLeuLysValThrThrThrAsnLysGlu 624
||| :::: :::: |||
1980 AAAAGATGCGCATCTACTATTAACTCAAGTGGATATAAACA..... 2022
624 LysLysLysAlaAspAlaTrpAspAlaGlnValThrValGlnAsnAsnSer 640
||||| :::: |||||
2023ACTGATAATGTTTAACTGTTGGTAAATAT... 2052
641 ValIleThrValGlyGlnGlyAlaLysAlaGlyLysValThr 657
2052 2052
657 rValValLeuAspGlyLysLeuIleThrThrIleSerPheLysValValA 674
2053GGTACTGCTGTCACTAAAGAGTGGCTTTGAACGTGTA 2089
674 spHraLaProThrAlaLysGlyLeuAlaValGlnPheThrSerThrSer 690

```

2090 AACCTGAGGAGCT.....GATGAGATTCGCGGTAATGAAGTAACT 2127
691 LeuLysGluValAlaIleProAsn...AlaAspLeuLysAlaIleLeuLys 706
:::||||:::|||||:::|||||:::|||||:::|||||:::|||||
2128 GTAAAGATCTCTACTGTAATGACGCTGATAAGAAAGTCCGCACTGTAAA 2177
706 nileLeuSerValAspGlyValProAlaIleThrAlaLysIleThrAla 723
:::||||:::|||||:::|||||:::|||||:::|||||
2178 AGATGTT.....GCACCGCAATTAAATGATGCGCGGA 2209
723 eAsnValGluPheValSerAlaAspThrAsnValAlaIleGluAsnGly 739
:::||||:::|||||:::|||||:::|||||:::|||||
2210 CTTTGTGGAAGAACAGACAAATTAACTAACTCTATTGATGAAGATAATCCT 2259
740 ThrValGluAlaLys.....GlyAlaThrSerIleTy 750
|||:::|||||:::|||||:::|||||:::|||||
2260 ACAGATTAACGCGCAAGATGACGCACTTAAAGCGGCGCATACCTTAACCT 2309
750 rVal.....LysAsnLeuThrValValLysAspGlyLysGluGlnLysV 765
:::|||||:::|||||:::|||||:::|||||:::|||||
2310 TAAAGCAGTTAAACCTGTAAGAGTTAAACGGATGAGAAA.....ATA 2353
765 AlGluPheAspLysAlaValGluValAlaValSerIleLysGluAlaLys 781
:::|||||:::|||||:::|||||:::|||||:::|||||
2354 TTACTTTTGACTTGGCGCAAAAAC.....CTTGAGGTGGAAGAACGCGGAA 2397

seq_name: /cgn2_6/ptodata/2/1na/6A.COMB.seq:US-08-685-467-3

seq_documentation_block:
  Sequence 3, Application US/08685467
  Patent No. 6060059
  GENERAL INFORMATION:
    APPLICANT: St. Gene III, Joseph W.
    APPLICANT: Barenkamp, Stephen J.
    TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
    NUMBER OF SEQUENCES: 6
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Flehr, Hobbach, Test, Albitton & Herbert
      STREET: Four Embarradero Center, Suite 3400
      CITY: San Francisco
      STATE: California
      COUNTRY: United States
      ZIP: 94111-4187
    COMPUTER READABLE FORM:
      MEDIUM TYPE: floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent In Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/685,467
      FILING DATE: 22-JUL-1996
      CLASSIFICATION: 424
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/409,995
      FILING DATE: 24-MAR-1995
      ATTORNEY/AGENT INFORMATION:
        NAME: Silva, Robin M.
        REGISTRATION NUMBER: 38,304
        REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (415) 781-1989
        TELEFAX: (415) 398-3249
        TELEX: 910 277299
    INFORMATION FOR SEQ ID NO: 3:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 5738 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: unknown
        TOPOLOGY: unknown
  MOLECULE TYPE: DNA (genomic)
  US-08-685-467-3

```

```

alignment_scores:
  Quality: 238.00      Length: 850
  Ratio: 0.580        Gaps: 42
  Percent Similarity: 48.235      Percent Identity: 22.824

alignment_block:
  alignment_09-754-947-1 x US-08-685-467-3 ..

Align seg 1/1 to: US-08-685-467-3 from: 1 to: 5738

12 HIRTPGILLeasPserILeasnTrILeuValGLuLysGLYValVally 28
   ::::::::::::::::::::
256 AATTGGCGAAR.....TATTCCACAATAAAGACTACTAAA 239
s.....GLYAsnAspLysGlyMetPheGluP 37
1 AGCCGAGACATTCACCTCCAAAGCCGGCGACACACTGAATAATTAACAAA 343
37 roGLYLySGluLeuThrArgAlaGluAlaIaIaThr.....MetMetAla 51
344 ANACCGATGAAGNCCATGCACGTAGCTTCACCTACCTGCGTGAAAAAA 393
52 GlnILeuAsnLeuProILeasPlysAspAlaLysProSerPheAlaAs 68
394 GACCTCACAGATCTGACACAGTGTCCACATGAAATAATATCGTTGGCGC 443
68 pSerGlnGlyInTrpLyrThrProPheILeAlaIaValGluLysAlaG 85
444 AAMCGCGCATTA.....GTTATATTTACCGAGTAA 478
85 LYValILeLys.....GLYThrGlyAsnGly...PheGluProAsnGly 98
479 ATGCGCTGAATTTGGCGAAACAGGTAAACGAAATGTTCAATTTGATGCT 528
99 .....LysILeAspArgValSerMetAlaSerLeuLeuValGlu 111
529 TTGATTCACATTTGCCGTGATGCGGTAAACGATACAGATGCTGTTAAGTTC 578
111 uAlaTrILysLeuAspThrLysValAsnGlyThrProAlaThrLysPheL 128
579 ATCAAGTTTTCACCTTAATGATGTTGAAAAACACAGACGTGCACACTGTTA 628
128 YAspLeuGluILeuthrLeuAsnTrpGlyLysGluLysAlaSnILeLeuVal 144
629 AAGATGTTTAAATGACAGTGTGAACATTAAAGCTCTTAAACT..... 672
145 GluLeuGlyLysSerValGlyThrGlyAspGlnTrpGluProLysLysTh 161
673 .....GCTGGAGGTAAATGTTGACAGTGTGATTT 701
161 rValThrLysAlaGluAlaIaGlnPheILeAlaLysThrAspLysGluP 178
702 AGTTCGCGCTTATATAATATGTTGAATTTATTCACGGCGATTAACACGCC 751
178 heGlyThrGluAlaIaLysValGluSerAlaLysAlaValThrTrnGln 194
752 TTGATGTTTGTTAACACTTAAGAAACNGTAAACAA..... 789
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211
790 ...ACCGAAGTAAATTCACACCGCAAAACCTGTGTTATCAAGAAAGAAAG 836
211 pILeLysVal...ThraSnLysAlaAsnAsnAsp.....LysValLeuV 225
837 CGGTAACTTATTTCTCGGAAAAAGACAAATTAACACCAAAATAAGTTACAA 866
225 alLysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
887 GTAAACAGCGCACATCATATACAGATGAGCGTAATGCGTTAAGCATCGCA 936
242 TyrSerAsnLeuAlaIaLysGlnThrLysValAspValAsnLysVal 258
   ::::::::::::::

```

```

937 AAAGCTGATGTGATGCT.....GTGAACAAGC 965
258 LGY.....LysThrGluValAla..... 264
    ||| |||||
966 TGGTTGCGACGACTTAACAACACTACTGCTAATGTCACAAATGCGCAGCTTCG 1015
265 .....ValGlySerLeuGluAlaLysThrIleGluMetAlaAspGlnThr 279
    ||| ||||| ||||| |||||
1016 CAACTGTTCCGCTCAGGCACAAATGTAACCTTTGAAAGTGGCGATGCTACA 1065
280 ValValAlaAspGluProThrAlaLeuGlnPheThrValLysAspGluAs 296
    : : : : :
1066 ACAGCGTCA.....GTAACCTAAAGATACTAA 1091
296 ngLYThrGluValAlaSerPro.....GluGlyIleG 307
    ||| : : : : :
1092 CGCGAATGCGATCACTGTTAAGTACGACGCCAAAGTTGGCGACGCTTGA 1141
307 LurPheValThrProAlaIleGluLysIleAsnAlaLysGlyGluIleThr 323
    : : : : :
1142 AATTT.....GATAGCGATAAAAAATC..... 1164
324 LeuAlaLysGlyThrSerThrThrValLysAlaValIleThrLysAspGln 340
    : : : : : |||||
1165 GTTCACATACGACGACCGACCTTACTGTACAGCT.....GG 1199
340 YLysValValAlaGluSerLysGlu.....ValLysValSerAlaG 354
    ||| |||||
1200 TAAAGTAGCTGAATAATGCTTAAGAAAGATGACAAAGAAACCTTGTTATG 1249
354 LurGlyAlaAlaValAlaSerIleSerAsn.....TrpThrValAlaGlu 368
    ||| ||||| : : : : :
1250 CAGCGCTTTGGTAAACAGCTTTAGGTAACTAAAGTGAAGCAAAAGCT 1299
369 GlnAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnLysVal 385
    : : : ||| : : : : :
1300 GAGCGTGATCTCATGCTGCGCTTGAAGCGATTTCAAAAGACCAAGAGCT 1349
385 LTrpGluGlyAspAsnAlaTrpValGluValGluLeuLysAspGlnPhe 402
    | ||||| : : : : :
1350 CAAAGCAGCGCAACGTA.....ACCTTTA 1375
402 snAlaValThrThrGlyLysValGluTrpGluSerLeuAsnThrGluVal 418
    : ||| : : ||||| : : : : :
1376 AACGGGCGCAAGAACTTAAGATGAACAGAGATGCGCAACTTTACTTAT 1425
419 AlaValAlaAspLysAlaThrGlyLysValThrVal..... 430
    : : : : : ||||| : : : : :
1426 TCACTCAGAGATGCTTTAAGCGGTTAAGCAGCATTTAGCTGTAC 1475
431 LeuSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
    : : : : : |||||
1476 AACCTAATGCGGAAATGATGCGAAACCGTCATCAACAAAGAGCGTTTAA 1525
445 .....SerLysGlyLysAlaLeuValSerHisThrValGlu 456
    : : : : : ||||| : : : : :
1526 CCATCAGCGCAGCAGTATGCGCGTACAGACAGTCAAAACACACATCAGC 1575
457 IleGluAlaPheAlaGlnLysAla...MetLysAspIleLysLeuGlu 472
    : : : ||||| |||||
1576 GTAAACCAAGATGCGATTAAACAGAGTAAAGCTATT..... 1614
472 ghrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLysA 489
    : ||||| ||||| : : : : :
1615 .ACTAATGTCGAGCTTAAAGCGTTATGACGAT..... 1650
489 laProValLeuAspGlnTrpGlyLysGluPheThrAlaProValThrVal 505
    : : : : :
1651 .....GCGAATTTT 1659
506 LysValLleuAspLysAspGlyLysGluLeuLysGluGln..... 518
    ||||| : : : : :
1660 GATGTTTAAATAACTGCAACTGATTTAAATAGACACAGCTGAAGATGC 1709

```

```

519 .....LysLeuGluAlaLysTrpValAlaAsnArgGlu...L 529
    : : : : : ||||| : : : : :
1710 TTATAAAGCTTTATTAATCTAAATGAAAAAATGAAACCAACCGT 1759
529 euValLleuAsnAlaIleGlyGlnGluAlaGlyAsnTrpThrVal..... 543
    ||||| : : : : :
1760 TGGTGACTGACAGCAGCGCGGCGACTGTAGCGATTACGTAATTGGGT 1809
544 ..ValLeuThrAlaLysSerGlyLysGluAla.....LysAlaThr 557
    ||||| : : : : : |||||
1810 TGGGTAGTATCAACCAAAAGCGTACGAAGAAGAACATCAAGTTAA 1859
557 rLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPheGluValA 574
    ||| ||||| |||||
1860 ACAAGCTGATGAAGTCTCTTTACCGGAGCC..... 1890
574 rgGlyLeuAspThrGluLeuAspLysTrpValThrGluLysGlnLys 590
    ||| |||||
1891 ..GGTGTGCTACCGTTACTTCCAAA.....TCTGAACCGTAA 1929
591 AsnAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuVal 607
    : : : ||||| |||||
1930 CATACGATTACCGTTAGTGTGCTGAACATAACCGGATGCCGCTTGA 1979
607 uLysGlyAlaGluAlaIleGluLeuLysValThrThrAsnLysGluG 624
    ||| : : : : : |||||
1980 AAAAGATGCGGATTAATTAAGCTCAAGTGATTAACAAAC..... 2022
624 LTrpGluValAspAlaThrAspAlaGlnValThrValGlnAsnSer 640
    ||||| : : : : : |||||
2023 .....ACTGATATGTTTAACTGTTGGTAAATAT... 2052
641 ValIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTrpLysValTh 657
2052 ..... 2052
657 rValValLeuAspGlyLysLeuIleThrThrHisSerPheLysValVal 674
    ||| ||||| : : : : :
2053 .....GTTACTGTCTCACTTAAAGTGGCTTGAACCTGTTA 2089
674 sPThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSer 690
    ||||| : : : : :
2090 AAACGTGAGCGACT.....GATGCAGATCGCGGTAAAGTAAC 2127
691 LeuLysGluValAlaProAsn...AlaAspLeuLysAlaAlaLeuLeuAs 706
    : : ||||| : : ||||| : : : : :
2128 GTAAAGATGCTACTCTTAATGACGCTGATAGAGAAAGTCGCAACTGTAA 2177
706 nIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrAla 723
    : : : : : ||||| : : : : :
2178 AGATGTT.....GCAACCGCAATTATATAGCGCGCA 2209
723 eraSnValGluPheValSerAlaAspThrAsnValAlaIleGluAsnGly 739
    : : ||||| : : ||||| : : : : :
2210 CTTTGTGTAAGACAGAAATTAACTACCTTAATGATGAGATATTCCT 2259
740 ThrValGlyAlaLys.....GlyAlaThrSerIleLeu 750
    || : : : : : |||||
2260 ACAGATAAAGCGCAAGATACGCACCTTAAAGCGGGATACCTTAACCTT 2309
750 rVal.....LysAsnLeuThrValValLysAspGlyLysGluGlnLysV 765
    : ||||| ||||| : : ||||| : : : : :
2310 TAAAGCAGGTAAAAACCTGAAGTTAAAGCTATGAGAAA.....AA 2353
765 aGluPheAspLysAlaValGlnValAlaValSerIleLysGluAlaLys 781
    : ||||| ||||| : : : : :
2354 TTACTTTGACTTGGCNAAAAAC.....CTTGAGGTGAAAACCTCGAAA 2397

```

```

seq_name: /cgn2_6/p10data/2/ina/5A_COMB.seq-US-08-038-682-8
seq_documentation_block:
; Sequence 8, Application US/08038682

```



```

439 LysValThrValLysAspSerLysGlyLysAlaLeuValSerHisThrVa 455
|||.....|
4425 AAAGTGAATTAAGTCACTTCGGTATATGTAATATTACAGCAGCGG 3474
|||.....|
445 LgluileGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluL 472
|||.....|
3475 CAATACACTTAAGTAACTAATATCAGTGTCAAGATCTAACAGTAACAG 3524
|||.....|
472 yStrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValys 488
|||.....|
3525 CGGATGCAGCAGCCTTGACAACTACAGCAGGCTCAACCATTAAGCCACA 3574
|||.....|
489 AlaProValLeuAspGlnIleGlyLysGluPheThrAlaProValThrVa 505
|||.....|
3575 ACAGGC.....AATGCAATATTACAC 3597
|||.....|
505 LysValLeuAspLysAspGlyLysGluLeuLysGluLysLeuGluA 522
|||.....|
3598 CAAACAGCTGATATCAACGCT.....AAAGTTGAAT 3629
|||.....|
522 LgluLysValAlaSerGluLeuValLeuAsnAlaAlaGlyGlnGluAla 538
|||.....|
3630 CCAGCTCCGCTCTGTACACTTGTTCAGACTGAGCAGCACTCTTCTCTGA 3679
|||.....|
539 GlysAsnTyr.....ThrValAlaLeuThrAlaLysSerGlyGluLys 552
|||.....|
3680 GGTATATTTCAGTACACTGTTACTTACTCTGCGGATAGCGGT..... 3724
|||.....|
552 sGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheS 569
|||.....|
3725 .....AAATTAACTCCACAGTAGTGTTCATCAATTAATGGACTAATA 3767
|||.....|
569 eLysPheGluValAlaGlyLeuAspThrGluLeuAspLysTyrValThr 585
|||.....|
3768 GT.....CTAAC 3775
|||.....|
586 GluLysAsnGlnLys.....AsnAlaMetThrVa 595
|||.....|
3776 ACCTCACACCAATCAGCGATATTGAGGTACAAATTTCTGGTAATACAGT 3825
|||.....|
595 LSerValLeuProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluA 612
|||.....|
3826 AAATGTT.....ACAGCAGCAGCTGGATTTTAACTTGAATTAAGTGC 3872
|||.....|
612 LalaGluLeuLys.....ValThrThrAsnLysGluLys 625
|||.....|
3873 AAGTTGAACGCGAAATGAGCTGCACCTTAACGTGGAATCAGGCAAA 3922
|||.....|
626 GluValAspAlaThrAspAlaGlnValThrValGlnAsn...AsnSerVa 641
|||.....|
3923 TTAACCCACCAACAGGCTCTAGCATTCCTCAGCAATGCTGCAGACAA 3972
|||.....|
641 LLeuThrValGlyGlnGlyAlaLysAlaGlyLysValThrValThrVa 658
|||.....|
3973 TCTTACACCCAAAGATGACAGATGCGCAGGAACATTAATGCTGCTAATG 4022
|||.....|
658 alValLeuAsp.....GlyLysLeuIleThrThrHisSerPheLysVal 672
|||.....|
4023 TGACGTTAAATTAACACAGGCACTTAACCTAAGTCAAGGATTCAGAA 4069
|||.....|
673 ValAspThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerTh 689
|||.....|
4070 .....ATTAAACGCAACCAAGTGG 4086
|||.....|
689 rSerLeuLysGluValAlaProAsnAlaAspLeuLysAlaAla..... 703
|||.....|
4087 TACCTTAACATCAATGCAAAAGATGCCAAATTAATGATGCTGCATCAG 4136
|||.....|
704 .....LeuLysAsnIleLeuSerValAspGlyValProAlaThr 716
|||.....|
4137 GTGACCGCAGCACTAGTAAATGCACTAACGCAAGTGGCTCTGTAAACGTG 4186
|||.....|

```

```

717 ThrAlaLysAlaThrAlaSerAsnValGluPheValSerAlaAsp..... 731
|||||.....|
4187 ACTGGGAAA...ACCTCAAGCAGCGTGAAAT...ATCACCGGGGATTTAA 4230
|||||.....|
732 .....ThrAsnValAlaAlaGluAsnGly.....ThrValG 742
|||||.....|
4231 CACATTAATGAGTTAAATATCATTCATTCGAAATAATGTAACCAACTGTC 4280
|||||.....|
742 LysAlaLysGlyAlaThrSerIleTyrValLysAsnLeuThrValValys 758
|||||.....|
4281 CTTTAAGAGGC...AAGCAATTCATGTGAAA.....TATATCCAA 4318
|||||.....|
759 AspGlyLysGluGlnLysValGluPheAspLysAlaValGlnValAlaVa 775
|||||.....|
4319 CCAGGTGAGCAAGCGTGAAGAGGTAATGAGCGGAACCGCTCTTGA 4368
|||||.....|
775 LSerIleLysGlu 779
|||||.....|
4369 GAGGTAAAGAT 4381
|||||.....|

```

```

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:us-08-302-832-8
seq_documentation_block:
; Sequence 8, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-8

```

```

alignment_scores:
Quality: 218.50      Length: 771
Ratio: 0.530         Gaps: 36

```


[illegible]


```

seq_name:/cgn2_6/plotdata/2/lna/5B_COMB.seq:US-08-469-880-8
seq_documentation_block:
; Sequence 8, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface ProteinS
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltace, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4702 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-8

alignment_scores:
Quality: 218.50 Length: 771
Ratio: 0.530 Gaps: 36
Percent Similarity: 53.437 Percent Identity: 22.438

alignment_block:
US-09-754-947-1 x US-08-469-880-8 ..

Align seg 1/1 to: US-08-469-880-8 from: 1 to: 4702

89 GYTHNGIYASngciYPheclupPoasnglyLysIIeasparYValSerme 105
|||||:.....:..:..:|||||:.....:.....:
2306 GGGGAAATTCAGAGCGTACGATTACGGGGCAATATCAATTCACCAATAA 2355
105 ValaserLeuLeuValGAlaIatYrLysLeuSprnHysValAsnglyT 122
|||||:.....:..:..:|||||:.....:.....:
2356 ACCAAATGTTACATTTCAAGACT.....:..:GACACCAGCAACAGCACA 2396
122 hPrcAlaLThrLysPheLysAspLysGluThrLeuAsnTrpGlyLysGlu 138
|||||:.....:..:..:|||||:.....:.....:

```



```

: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 4702 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-728-470-8

alignment_scores:
  Quality: 218.50      Length: 771
  Ratio: 0.530        Gaps: 36
  Percent Similarity: 53.437  Percent Identity: 22.438

alignment_block:
US-09-754-947-1 x US-08-728-470-8 ..
Align seg 1/1 to: US-08-728-470-8 from: 1 to: 4702

89 GlyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArgValSerIle 105
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2306 GGGGAAATTTCAAGCAGTAGCATTTACGGGCAATATCAATATCCACCAATTA 2355
105 ValSerLeuLeuValGluAlaIleTyrLysLeuAspThrLysValAsnGlyT 122
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2356 ACCAATCTTACATTACAGCT.....GACACACAGCAACAGCAACA 2396
122 hPrcAlaThrLysPheLysAspLeuGluThrLeuAsnThrPglYsglu 138
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2397 CAGGCTTCAGAGAA.....AGAACTCAACTCTGGC..... 2428
139 LysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThrGlyAspGl 155
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2429 .....AATATATCTGTGGGCAATTTTAAGCTTA...ACTGGTGCAAA 2469
155 nTrpGluProLysLysThrValThrLysAlaGluAlaIleGlnPheIleA 172
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2470 TGCAAACATTTGCGGCAATCTTCTATTTCACAA..... 2503
172 ValLysThrAspLysGlnPheGlyThrGluAlaIleLysValGluSerAla 188
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2504 .....GATTCCACATTTAAAGAGAACGACGACCAACTAAACATC 2545
189 LysAlaValThrThrGlnLys.....ValGluValLysPheSerLysAl 203
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2546 ACCGGCACCTTTACCAACAACGCGTACCGCCACATTATATATAAACAAAGG 2595
203 ValGluLysLeuThrLysGluAspIleLys.....V 214
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2596 ACTGGTAAAACTCCAAAGCGGATATTATCAATAAGGTGTTAAATATCA 2645
214 AlThrAsnLysAlaAsnAsnAspLysValLeuValLys..... 226
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2646 CTACTAACGCCCTCAGGCACTCAAAAAACCATTTAAACGGAATATAACT 2695
227 .....GluValThrLeuSerGluAspLysArgSerAlaThrVa 239
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2696 AACGAAAAAGGCGACTTAACATCAAGATTATTAAGCCGACGCCGAAT 2745
239 GluLeuTyrSerAsnLeuAlaIleLysGln...ThrTyrThrValAspV 255
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2746 CCAAAATTCGGCGCATATCTCTCAAAAAAGAACGACATCTCACAAATTCT 2795
255 AlaAsnLysValGly.....LysThrGluValAla 264
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2796 CTGATTAAGTAATAATATTACCAATTCAGATTAAACAAAGCAGCGCTTGA 2845
265 ValGlySerLeuGluAlaIleLysThrIleGluMetAlaAspGlnThrValVa 281
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2846 GGGGGCCCTTCTGATTCAAGTGAGCGAAGAAATGCTAACCTAACATTAATCA 2895
```

```

281 lAlaAspGlu...ProThrAlaLeuGlnPheThrValLysAspGluAsnG 297
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2896 AACCAAGAGGTTAAATTTGGCAGAGACCTAAATATATTAGGCTTTAATA 2945
297 lYThrGluValValSerProGluGlyIleGluPheValThrProAlaAla 313
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2946 AAGCAGAAATTCACGCTAAAAATGGCAGATTTAACTATTGGCAATGCT 2995
314 GluLysIleAsnAlaLysGly...GluIleThrLeuAlaLysGlyThrSe 329
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
2996 AGCGGTGTAAATGCTGATGCTAAAAAGTAGCTTTTGACACAGGTTAAAGA 3045
329 rThrThrValLysAlaValIleTyrLysAspGlyLysValAlaIleGlu 346
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3046 TTCAAAAATCTCGACT.....GACGCTCACAATGTATACACTTA 3083
346 eLysGluValLysValSerAlaGluGlyAlaAlaValAlaSerIleSer 362
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3084 ATAGCGAAGTGAAGAACGTCT...AATGTATGATGCAATGCTGTATGAT 3130
363 AsnTrpThrValAlaIleGluIleAsnLysAlaAspPheThrSerLysAsp 379
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3131 AACAGCACCGGTTTAACCATTTCCGCAAAAAGATGTACGTTAAACAAATA 3180
379 eLysGluAsnAsnLysValIleTyrGluGlyAspAsnAlaTyrValGlnVal 396
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3181 CGTTACCTCCCAAGACATTAATATATCTCTCCGACAGCAGGAATGTAA 3230
396 lLeuLysAsp.....GlnPheAsnAlaValThrThrGlyLysValGlu 410
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3231 CAACCAAAAGAACGACCAACTATCAATGCA...ACCACAGCAGCGTGGA 3277
411 TyrGluSerLeuAsn..... 415
3278 GTAACTGCTCAAAATGTATCAATTAAGCAACATTACTCCCAAAATGT 3327
416 .....ThrGluValAlaValAla 422
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3328 AACAGTGACAGACAGCAAAATCTGTATCCACAGAGAAATGCTGCATTA 3377
422 sPLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaProVal 438
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3378 ATGCAACGCGCGCACAGTAAACATT...AGTACAAAACAGGGGATAT 3424
439 LysValThrValLysAspSerLysGlyLysAlaLeuValSerHisThrVa 455
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3425 AAAAGTGCAATTGAATCAACTTCGCGTATGTAAATATTACACGCGAGCG 3474
455 lGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluL 472
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3475 CAATACACTTAAGTAAGTAATATATCATCTGCTCAAGATGTAAACAGTAAC 3524
472 ystThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLys 488
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3525 CGGATGCAGAGACCTTGACACATACAGCAGGCTCAACCTTGTGCGACA 3574
489 AlaProValLeuAspGlnIleTyrGlyLysGluPheThrAlaProValThrVa 505
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3575 ACAGGC.....AATGCCAAATATTACAAAC 3597
505 lLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGluVa 522
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3598 CAAACAGGTGATATCAACGGT.....AAAGTTGAAT 3629
522 lAlysTyrValAsnArgGluLeuValLeuAsnAlaIleGlyGlnGluAla 538
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3630 CCAGCTCCGGCTCTGTAAACATTTGTGCAACGTGACCAACTCTCTGTA 3679
539 GlyAsnTyr.....ThrValValLeuThrAlaLysSerGlyGlyLys 552
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
3680 GSTAATATTTCAGTAACTGTTACTATTACTGCGGATAGCGGT..... 3724
```


[illegible]


```

339 rThrValLysAlaValTyrLysLysAspGlyLysValAlaGluS 346
    ::::: ::::: ||||| :::::
3261 TTCAAATACTCGACT.....GAGGTCACATGTAACACTAA 3298
    ::::: :::::
346 eRLysGluValLysValSerAlaGluGlyAlaAlaValAlaSerL 362
    ::::: ||||| ::::: ::::: :::::
3299 ATAGGGAAGTGAACGCTCT...AATGGTAGTACAAATGCTGGTAT 3345
    ::::: ||||| :::::
363 AsnTrpThrValAlaGluLysLysAlaAspPheThrSerLysAsp 379
    ||| ||| ||| |||
3346 AACAGCACCAGGTTAACCATTTCCGCAAAAGTGAACGGTAACATA 3395
    ::::: ||||| :::::
379 eLysGlnAsnAsnLysValTyrGluGlyAspAsnAlaTyrValG 396
    ::::: ||||| :::::
3396 CGTTACCTCCCAAGACAAATGAATATCTCTCCGACAGCAAGAA 3445
    ::::: ||||| :::::
396 LuLeuLysAsp.....GlnPheAsnAlaValThrThrGlyLys 410
    ||||| ::::: |||||
3446 CAACCAAGAAAGGCACAACTATCAATGCA...ACCACAGGCGG 3492
    ::::: |||||
411 TyrGluSerLeuAsn..... 415
    ::::: |||
3493 GTAACCTGCTCAAAATGTACATTAAGGCACATTACCTCGCAAA 3542
    ::::: |||
416 .....ThrGluValAlaValAla 422
    ::::: |||||
3543 AACACTGCAGCAGCAAAATCTGTACACAGAGAAATGCTGCATTA 3592
    ::::: |||||
422 sPLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAla 438
    ::::: |||||
3593 ATGCACACGAGCGCACAGTAACATT...AGTACAAAACAGAGGAT 3639
    ::::: |||||
439 LysValThrValLysAspSerLysGlyLysAlaLeuValSerHis 455
    ::::: |||||
3640 AAGGTGGAATGATCAACTCCGGTAAATGTAATTAACAGCGAGCG 3689
    ::::: |||||
455 lGluLeuAlaPheAlaGlnLysAlaMetLysAspIleLysLeuG 472
    ::::: |||||
3690 CAATACACTTAAGTAAGTAATATCATCTGTCACAGATGAACAG 3739
    ::::: |||||
472 ySThrAsnValAlaLeuSerThrLysAspValThrAspLeuLys 488
    ::::: |||||
3740 CGGATGCAGAGACCTTGACAACTACAGAGCTCAACCTTAGTG 3789
    ::::: |||||
489 AlaProValLeuAspGlnTyrGlyLysGluPheThrAlaProVal 505
    ::::: |||||
3790 ACAGCG.....AATGCAAAATATTACAC 3812
    ::::: |||||
505 lLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLys 522
    ::::: |||||
3813 CAAAACAGCTGATATCAACGCT.....AAAGTTGAAT 3844
    ::::: |||||
522 lAlaLysTyrValAsnArgLysLeuValLeuAsnAlaAlaGlyG 538
    ::::: |||||
3845 CCAGCTCCGGCTCTGTAACTTTGCAACTGGAGCAACTTCTGCT 3894
    ::::: |||||
539 GlyAsnTyr.....ThrValValLeuThrAlaLysSerGlyGlu 552
    ||||| ::::: |||||
3895 GGTAAATATTTCAGTAACACTCTTACTATTACGCGAGAGCGGT... 3939
    ||||| ::::: |||||
552 sGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAl 569
    ||| ||| :::::
3940 .....AAATTAACTCCACAGTGTCTACAAATTAAATGGGACT 3982
    ||| ||| :::::
569 eRLysPheGluValArgGlyLeuAspThrGluLeuAspLysTyr 585
    ||| ||| :::::
3983 GT.....GTAAAC 3990
    |||
586 GluGluAsnGlnLys.....AsnAlaMetThrVal 595
    ::::: |||||
3991 ACCTCAACCCATCAAGCGATATTGACAGGTACAATTTCTGGTA 4040
    ::::: |||||
595 lSerValLeuProValAlaPheAlaAsnGlyLeuValLeuLysG 612

```

```

4041 AAATGTT...ACAGCAACGACTGTGATTTAATCTATTGGAAATAG 4087
    ::::: |||||
612 lAlaGluLeuLys.....ValThrThrAsnLysGluGlyLys 625
    ::::: |||||
4088 AAGTTGAAGCCAAAATGAGAGCTGCACACCTTAATCTGTAATCA 4137
    ::::: |||||
626 GluValAspAlaThrAspAlaGlnValThrValGlnAsn...As 641
    ::::: |||||
4138 TTAACCAACCCAAACAGGCTGTAGCATTAACCTCAACCAATGTC 4187
    ::::: |||||
641 lIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysVal 658
    ::::: |||||
4188 TCTTACAGCCAAAGATAGCAGATGTCAGAGAAACATTAATGCT 4237
    ::::: |||||
658 aIValLeuAsp.....GlyLysLeuIleThrThrHisSerPhe 672
    ::::: |||||
4238 TCACGTTAAATACACAGGCACTTAATCTACTACACAGGATTA 4284
    ::::: |||||
673 ValAspThrAlaProThrAlaLysGlyLeuAlaValGluPhe 689
    ::::: |||||
4285 .....ATTAACGCAACCAAGTGG 4301
    ::::: |||||
689 lSerLeuLysGluValAlaProAsnAlaAspLeuLysAlaAla... 703
    ::::: |||||
4302 TACCTTAACATCAATGCAAAAGATGCCAAATTAAGATGGTGTG 4351
    ::::: |||||
704 .....LeuLeuAsnIleLeuSerValAspGlyValProAla 716
    ::::: |||||
4352 GTGACCGCAGCATAGTAATGCAACTAACGCAAGGCGCTGTAG 4401
    ::::: |||||
717 ThrAlaLysAlaThrAlaSerAsnValGluPheValSerAlaAs 731
    ::::: |||||
4402 ACTGGGAAA...ACCTCAAGCAGCGTGAAT...ATCACCGGG 4445
    ::::: |||||
732 .....ThrAsnValAlaGluAsnGly.....ThrValG 742
    ::::: |||||
4446 CACAATTAATGGTTAAATATATATTCGGAAAATGTAACAACTG 4495
    ::::: |||||
742 lValAlaLysGlyAlaThrSerIleTyrValLysAsnLeuThr 758
    ::::: |||||
4496 GCTTAAGAGC...AAGGAATTGATGTCAA.....TATATCCA 4533
    ::::: |||||
759 AspGlyLysGluGlnLysValGluPheAspLysAlaValAla 775
    ::::: |||||
4534 CCAGGTGTAGCAAGGTAGAAAGGTAATTTGAAGGAAACGCTCT 4583
    ::::: |||||
775 lSerIleLysGlu 779
    ::::: |||||
4584 GAAGTAAAGAT 4596

```



```
/gene="sap"
/Note="surface layer homology motif, SLH1"
misc_feature      866..1042
                  /gene="sap"
                  /Note="surface layer homology motif, SLH2"
misc_feature      1043..1216
                  /gene="sap"
                  /Note="surface layer homology motif, SLH3"
BASE COUNT      1360 a      471 c      645 g      986 t
ORIGIN

alignment_scores:
    Quality: 3847.00      Length: 785
    Ratio: 4.919          Gaps: 0
    Percent Similarity: 99.618      Percent Identity: 98.981

alignment_block:
US-09-754-947-1 x BASAPG ..

Align seg 1/1 to: BASAPG from: 1 to: 3462

1 AlaGlyLysThrPheProAspValProAlaAspHisTrpGlyIleAspSe 17
|||||
674 GCAGGTAAACAATTCACAGAGCTTCCTGTCATCAGTGGGAAATTCATTC 723
|||||
17 rIleAspTrpLeuValGluValAlaValLysGluValAsnAspLysGly 34
|||||
724 TATTACTACTAGTAGAAGAAAGGCGCAGTTAAAGTTAAAGTAAACGAA 773
|||||
34 eLpGluProGlyLysGluLeuThrArgAlaGluAlaAlaThrMetMet 50
|||||
774 TGTTCAGCGCTGGAAGAAAGATTACTCGTCGAGAGCAGCTCAATGATG 823
|||||
51 AlaGlnIleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAl 67
|||||
824 GCTCAAACTCTAACTTACCAATCGATTAAGAGTCTAAACCATCTTTCGC 873
|||||
67 AAspSerGlnGlyGlnTrpTrpThrProPheIleAlaAlaValGluLysA 84
|||||
874 TGACCTCTCAGAGCCCAATGTACACTCCATCTCCAGCTGTAGAAAGAA 923
|||||
84 IaGlyValIleLysGlyThrGlyAsnGlyPheGluProAsnGlyLysIle 100
|||||
924 CTGGCGGTATTAAAGGTACAGGAAACGGCTTGAGCCAAACGAAAAATC 973
|||||
101 AAspArgValIserMetAlaSerIleuValGluAlaIleTrpLysLeuAsp 117
|||||
974 GACCCGCTTCTATGCGATCTCTCTGTAGAAAGCTTACAAATTAGATAC 1023
|||||
117 rLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluThrLeuA 134
|||||
1024 TAAAGTAAACGCTACTCCAGCAACTAAATTTCAAGATTTAGAAACATTAA 1073
|||||
134 sNrTrpGlyLysGluLysAlaAsnIleLeuValGluLeuGlyIleSerVal 150
|||||
1074 ACTGGGGTAAAGAAAGCTAAACATCTTAGTGAATTAGGAATCTCTGTT 1123
|||||
151 GlyThrGlyAspGlnTrpGluProLysLysThrValIleThrLysAlaGlu 167
|||||
1124 GGTACTCGTGATCAATGGAGCCTTAAGAAACTGTAAAGCAGCAAGC 1173
|||||
167 AlaGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAlaAlaL 184
|||||
1174 TGCTCAATTCATTGCTTAAGACTGACAGCAGTTGCTACAGAAAGCAGCA 1223
|||||
184 ySValGlnSerAlaLysAlaValThrGlnLysValGluValLysPhe 200
|||||
1224 AAGTGAATCTGCAAAAGCTGTACAACTCAAAAGTAAAGTAAATTC 1273
|||||
201 SerLysAlaValGluLysLeuThrLysGluAspIleLysValIleThrAsn 217
|||||
1274 ACCAAAGCTGTGAAAAATTAACTAAAGAGATTCAAAGTAACTAAACAA 1323
```

```
217 sAlaAsnAsnAspLysValLeuValLysGluValThrLeuSerGluAspL 234
|||||
1334 AGCTAACACAGTAAAGTACTAGTTAAAGAGCTACTTATCAGAAAGTAA 1373
|||||
234 ySArgSerAlaThrValGluLeuTrpSerAsnLeuAlaLysGlnThr 250
|||||
1374 AAAAATCTGTACAGTTCAATTAATTAATTAATTAATTAATTAATTAAT 1423
|||||
251 TyrThrValAspValAsnLysValGlyLysThrGluValAlaValGlySe 267
|||||
1424 TACACTGTAGATGTAAACAAAGTTGTAAACAGAGTACGTAGGTTC 1473
|||||
267 rLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAspG 284
|||||
1474 TTTAGAACCAAAACAATCGAAATGCTGACCAAAACAATGTAGCTGATG 1523
|||||
284 lUpProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluVal 300
|||||
1524 AGCCAAACAGCATTTCAATTCACAGTTAAAGATGAAGAAACGTACTGA 1573
|||||
301 ValSerProGluGlyIleGluPheValThrProAlaAlaGluLysIleAs 317
|||||
1574 GTTTCACCAAGAGGTATTGAAATTGTAAAGCCAGCTGCAGAAAAATTAA 1623
|||||
317 nAlaLysGlyGluIleThrLeuAlaLysGlyThrSerThrValLysA 334
|||||
1624 TGCAAAAGGTGAATCACTTTAGCAAAAGGTCACTCACTACTGTTAAAG 1673
|||||
334 lAlaValTrpLysAspGlyLysValValAlaGluSerLysGluValLys 350
|||||
1674 CTGTTTATAAAAAGACGTTAAAGTAGTAGCTGAAGAGTAAACAAATAAA 1723
|||||
351 ValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTrpThrValAl 367
|||||
1724 GTTTCGTGGAAGGTGCTGCACTTACCTCAATCTTACTGACAGTTGC 1773
|||||
367 aGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAsn 384
|||||
1774 AGAACAAATATAAGCTGACTTACTTAAAGATTCAACCAACAAACATA 1823
|||||
384 ySValTrpGluGlyLysAsnAlaTrpValGlnValGluLeuLysAspGln 400
|||||
1824 AAGTTTACCAAGGCGACAAACGCTTACGTTCAAGTGAATTCAGAAATCA 1873
|||||
401 PheAsnAlaValThrThrGlyLysValGluTrpGluSerLeuAsnThrGl 417
|||||
1874 TTTAACGCAAGTAAACACTGGAAGATGTAATGAGTGTAAACACAGA 1923
|||||
417 uValAlaValAlaAspLysAlaThrGlyLysValThrValIleSerAlaG 434
|||||
1924 AGTTGCTGTAGTAGATTAAGCTACGTGTAAGTAACTGTATTAATCTGC 1973
|||||
434 lLysAlaProValLysValThrValLysAspSerLysGlyLysAlaLeu 450
|||||
1974 GAAAGACACCAATAAAGTAACTGTAAAGATTCAGAAAGTAAAGACTT 2023
|||||
451 ValSerHisThrValGluIleGluAlaPheAlaGlnLysAlaMetLysAs 467
|||||
2024 GTTTCAAAAACAGTTGAATTTGAGCTTCCGCTCAAAAAGCAATGAAGA 2073
|||||
467 pLleLysLeuGluLysThrAsnValAlaLeuSerThrLysAspValThrA 484
|||||
2074 AATTTAAATTAGAAAAAACTAACGTAGCGCTTCTCAAAAAGATGTAAAC 2123
|||||
484 sPLeuLysValLysAlaProValIleuAspGlnTrpGlyLysGluPheThr 500
|||||
2124 ATTTTAAAGTAAAGAGCTCCAGTACTAGATCAATACGTTAAAGATTTAA 2173
|||||
501 AlaProValThrValLysValLeuAspLysAspGlyLysGluLeuLysGl 517
|||||
2174 GCTCCTGTAAACAGTGAAGTACTGATTAAGATGTGAAGATTTAAAGA 2223
```

517 uGlnlySLeuGluAlaIAsnTYrValAsnArgGluLeuValLeuAsnAla 534
|||||
2224 ACAAAATTTGAAGCTTAATATGTAAGCAAAAGATTGATCTGTAATCAG 2273
534 IagIIngluAlaGlyAsnTYrThrValValLeuThrAlaIAsnSerGly 550
|||||
2274 CAGCTCAAGAAAGCTGTAATTATACACTTGTATTAACTGCAGAAATCTGCT 2323
551 GluIySGluAlaIAsnTYrThrValLeuGluLeuIySAlaProGlyAl 567
|||||
2324 GAAAAGAACCAAAAGCTACATCTAGCTCTGAAATTAAGACTCCAGGCTGC 2373
567 aPheSerLysPheGluValArgGlyLeuAspThrGluLeuAspLysTYr 584
|||||
2374 ATTCTCTAAATTTGAAGCTCGTGGTTTGAAGAAAAGAAATTGATTAATG 2423
584 aThrGluGluAsnGluIySAlaMetThrValSerValLeuProVal 600
|||||
2424 TTACTGAGGAAACCAAAAGATGCATGACTGTTTCAGTTCTTCTCTGTA 2473
601 AspAlaAsnGlyLeuValLeuIySGlyAlaGluAlaIagIuLeuIySva 617
|||||
2474 GATGCAATAGATGATGATTAATAAGGTGCAGAAAGCAGCTGAATAAAGT 2523
617 IThrThrThrAsnIySGluGlyIySGluValAspAlaThrAspAlaGln 634
|||||
2524 AACACACACAAACCAAGAGTAAAGAGTAAGACGCAACTGATGACACAG 2573
634 aThrValGlnAsnAsnSerValIleThrValGlyIngluAlaIySAla 650
|||||
2574 TTACTGACCAAAATTAACAGTGTATTACTGTGTGTCAGAGTGCAGAAAGCT 2623
651 GlyIuThrTYrIySValIThrValValLeuAspGlyIySLeuIleThrTh 667
|||||
2624 GGGAACCTATAAAGTAACAGTGTACTGATGATGTAATTAATCACACAC 2673
667 rHisSerPheIySValValAspThrAlaProThrAlaIySGlyLeuAla 684
|||||
2674 TCATCTCTTAAAGTGTGATACAGCACCAACTGCTAAAGGATTACAG 2723
684 aGluPheThrSerThrSerLeuIySGluValAlaProAsnAlaAspLeu 700
|||||
2724 TAGAATTACAGCACATCTCTTAAAGAGTACCTCCAAATGCTGATTTA 2773
701 LysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValProAlaThrTh 717
|||||
2774 AAGCTCAGCTTTAAATATCTTATCTGTGATGCTGACCTGGACACTAC 2823
717 rAlaIySAlaThrAlaSerAsnValGluPheValSerAlaAspThrAsn 734
|||||
2824 AGCAAAAGCAACAGTTCTATATGTAATTTGTTTCTGCTGACACAATG 2873
734 aValAlaIagIuAsnGlyThrValGlyValAlaIySGlyAlaThrSerIleTYr 750
|||||
2874 TTGTAGCTGAAATGTAACAGTTGTCAGAAAGGTGCAGAACCTCTATCTAT 2923
751 ValIySAsnLeuThrValValIySAspGlyIySGluIySGluIySValGluPh 767
|||||
2924 GTGAAAAAAGCTGACAGTTGTAAGAGATGGAAGCAAAAGAGTAAGATT 2973
767 eAspLysAlaValGlnValAlaValSerIleIySGluAlaIySProAla 784
|||||
2974 TGATTAAGCTGTACAAAGTTCAGTTCTATTAAGAAAGCAAAAGCTGCAA 3023
784 hIyS 785
|||||
3024 CAAAA 3028
seq_name: gb_ba:BRH012290
seq_documentation_block: 4082 bp DNA BCT 25-OCT-1998
LOCUS BRH012290
DEFINITION Bacillus thuringiensis etc gene.

ACCESSION AU012290
VERSION AU012290.1 GI:3805794
KEYWORDS ctc gene; CTC protein; S-layer protein.
SOURCE Bacillus thuringiensis.
ORGANISM Bacillus thuringiensis
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
1 (bases 1 to 4082)
REFERENCE
AUTHORS Sun, M.
TITLE Direct Submission
JOURNAL Submitted (22-Oct-1998) Sun M., Department of Microbiology,
Huazhong Agricultural University, Department of Microbiology,
Huazhong Agricultural University, Wuhan, Hubei 430070, CHINA
2 (bases 1 to 4082)
REFERENCE
AUTHORS Sun, M. and Yu, Z.
TITLE S-layer protein gene of Bacillus thuringiensis CTC
JOURNAL Unpublished
FEATURES
source location/Qualifiers
1. 4082
/organism="Bacillus thuringiensis"
/sub_species="Finlilimus"
/strain="CTC"
/db_xref="taxon:1428"
/note="Flagellar serotype H2"
718. 3168
/gene="ctc"
718. 3168
/gene="ctc"
/function="S-layer protein"
/codon_start=1
/transl_table=11
/product="CTC protein"
/protein_id="CAA09981.1"
/db_xref="GI:3805795"
/translation="MAKINSYKKVIAGTMTAAVGVSPYAAGKSPDVPADHMG
DSINLVKGAVTGNDKGMPEPGEKLTBAEATMAOILNLPIDKAPSPADSGOM
YTPPIAAVEKAGVINGTNGPBNCKIDRYSALSLVAKLIDTPVNGTPTAKPESL
TLNMGKEKANIIVELGISVGTADMEPKTYTKAAQOFIAKTTQKSAATVELSNLA
KAVTTQKVEKFSKAVEVITKEDVLANKANDKVLVDKLSBKSAATVELSNLA
AKQTYTVNKKVGEVTVGSLKAKTIEMADTVVADRPALAKTVKDENGEVVSPA
GIEFVTPAAEKINAKGEITLAKGSTYTKAVYKKGKVAESKEVVAEGTAVASIS
NMVTAERADFTSKDFKONDKYEGDNNVSVQELKDOHNNVNNVNAKAESELENTEAV
VDKATGKTVLSAGKAPYKYVYKDSKKEIYSKYEIARFQKAMKEIKLEETNVALS
TKDVDFKVKAPVLDQYGERAPAVEVYVLDQGELEBOKLVAKYENKELVLANHG
BAGKYTELTAASKGEVSKLALBLKAPGVSKDVGLENELEDKVYTEENKKNEMV
VSLPVDANGIVLREKEATLKTVTDDQKVDAISQVAVNDAAGTITVGNENKAG
EYKVTVAADGKLITTHSEFKVVDTPAPAKKLAVDPTSLINEVAGSELEKTAALNLS
VDGVPATAGATVTDVKFVVSADNTNVSEETAKFGTKSIFVKELTVKKGQPKVE
LDKPYRVDSIYEKVEK"
BASE COUNT 1574 a 543 c 765 g 1200 t
ORIGIN
alignment_scores:
Quality: 3304.50 Length: 785
Ratio: 4.424 Gaps: 5
Percent Similarity: 95.159 Percent Identity: 86.497
alignment_block:
US-09-754-947-1 x BRH012290 ..
Align seg 1/1 to: BRH012290 from: 1 to: 4082
1 AlagIySLyThrPheProAspValProAlaAspHisTrpGlyIleAsp 17
|||||
805 GCAGGAAATCATTTCCAGACGTTCCAGCTGATCAGTGGGAATTGATTC 854
17 rIleAsnTYrLeuValGluIySGlyAlaValIySGlyAsnAspIySGly 34
|||||
855 TATTAAGTACTTAGTAAAGGCGCAGTTACAGATTAAGATAAGGAA 904
34 ePheGluProGlyIySGluLeuThrArgAlaGluAlaIatThrMetMc 50
|||||

```

905 TGTTCGAGCCTGGAAAGAAATTAACTGCTGACAGACGCTACATGATG 954
51 AlaGlnIleLeuAñLeuProIleAspLysAspAlaLysProSerPheAl 67
955 GCTCAAAATCTTAAACTTACCAATCGATTAAGATGCTTAACCATCTTCCG 1004
67 aAspSerGlnGlyGlnTrpTyrThrProPheIleAlaLysAlaGluLysA 84
1005 TGACTCTCAAGGCCAAATGGTACACTCCATTCATCCGACGCTGTAGAAAAAG 1054
84 laGlyAlaIleLysGlyThrGlyAsnGlyPheGluProAsnGlyLysIle 100
1055 CTGGCCTTATTAAAGGTACAGCAAGCGCTTTGAGCCAAACGAAAAATTC 1104
101 AspArgValSerMetLAserLeuLeuValGluAlaTyrLysLeuAspTh 117
1105 GACCGCGTTCAATGGCATCTCTCTGTGAAAGCTTACAATTATGATAC 1154
117 rLysValAsnGlyThrProAlaThrLysPheLysAspLeuThrLeuA 134
1155 TAAAGTAAACGGTACTCCAGCACTAAATTCAAGATTATGAACATTAA 1204
134 snTrpGlyLysGluLysAlaAsnIleLeuValGluLeuGlyIleSerVal 150
1205 ACTGGGCTAAGAAAAAGCTAACATCCTAGTTGAATTAGGAATCTCTGT 1254
151 Gly...ThrGlyAspGlnTrpGluProLysLysThrValThrLysAlaG 166
1255 GGTACTACAGCTGATAGTGGGACCTTACGAAACCTGACTAAAGCAGA 1304
166 uAlaAlaGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAla 183
1305 AGCTGCTCAATTTATGTCTAAGACTGACAAACAGTTGCGTAGCAAGTAG 1354
183 laLysValGluSerAlaLysAlaValThrThrGlnLysValGluValLys 199
1355 CAAAAGTTGATCTGCTAAAGCAGTTACACTCAAAAGTAGAAGTTAA 1404
200 PheSerLysAlaValGluLysLeuThrLysGluAspIleLysValThrAs 216
1405 TTCAGCAAACTGCTGAAAAATTAACTAAAGAAAGACGTTAAATTGACTA 1454
216 nLysAlaAsnAsnAspLysValLeuValLysGluValThrLeuSerGlu 233
1455 CAAAGCTAACACAGATTAAGTATTAAAGATCTAAATTAATATGACGAA 1504
233 sPLysArgSerAlaThrValGluLeuTyrSerAsnLeuAlaLysGln 249
1505 ACAAAAATCTGCTACAGTTGAATTATAGTAAGTGTAGCTGCTAAACAA 1554
250 ThrTyrThrValAspValAsnLysValGlyLysThrGluValAlaValG 266
1555 ACTTACAGCTGATGATTAATTAAGTGTGTAAGTAGAGGTAACTGTAGG 1604
266 ySerLeuGlnAlaLysThrIleGluMetAlaAspGlnThrValValAla 283
1605 TTCCTTAACCAAAACAAATTTGAAATGGCTGACCAACAGCTGTACTCG 1654
283 sPGluProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGlu 299
1655 ATGAGCCACATGCAATTAATACAGCTTAAAGATGAAGTGAACGCTAGCA 1704
300 ValValSerProGluGlyIleGluPheValThrProAlaIleGluLys 316
1705 GTTGTCTTACCAGCGGCTATTGAATTGTAACTCCAGCTGCGAATAAAT 1754
316 eaSnAlaLysGlyGluIleThrLeuAlaLysGlyThrSerThrThrValL 333
1755 TAATGCAAAAGGTGAATTAATTAAGTAAAGGTACTTCACTACAGTAA 1804
333 ysaLysValTyrLysLysAspGlyLysValValAlaGluSerLysGluVal 349
1805 AACCTGTTATATAAAAGACGGTAGTGTAGTGAAGTAAAGAAAGTA 1854
350 LysValSerAlaGlnGluAlaAlaValAlaSerIleSerAsnTrpThrVal 366
1855 AAAGTTTCTGCTGAGAGTACTGACAGTAGCTTCAATCTTAAGTGGACAGT 1904
366 laLaGlnGluAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnA 383
1905 AGCAGCGAAG...AAAGTGATTTCACTTCTTAAAGACTTCAAAACAAAC 1951
383 snLysValTyrGlnGlyAspAsnAlaTyrValGlnValGluLeuLysAsp 399
1952 ATTAAGTTATGAAAGGTGACACGGTTTCTGTCACAGTGAATTAAGAC 2001
400 GlnPheAsnAlaValThrThrGly...LysValGluTyrGluSerLeuAs 415
2002 CATTTCAATTAAGCTGTAAATATACGTTAAAGCTGAATATGATCATTA 2051
415 nThrGluValAlaValAlaAspLysAlaThrGlyLysValThrValLeu 432
2052 TACAGAGGTAGCTGTAGATTAAGTACGTACGTAAAGTAACTGATATAT 2101
432 eAlaGlyLysAlaProValLysValThrValLysAspSerLysGlyLys 448
2102 CTGCAAGAAAACACCACTTAAGAGTACTGTTAAAGATTCAAAAGGTAA 2151
449 AlaLeuValSerHisThrValGluIleGluAlaPheAlaGlnLysAla 465
2152 GAACCTGTTCAAAAACGATTGAATTAAGCTTTTCTTCCATAAAACGCA 2201
465 LysAspIleLysLeuGlnLysThrAsnValAlaLeuSerThrLysAsp 482
2202 GAAGAATTAATTAAGTAAAGTAAAGCTCCAGTACATCAATATGCTAAAG 2251
482 aThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGlu 498
2252 TAACAGATTTCAAAAGTAAAGCTCCAGTACATCAATATGCTAAAGAG 2301
499 PheThrAlaProValThrValLysValLeuAspLysAspGlyLysGlu 515
2302 TTGCGAGCTCTCTAGAAAGTGAAGTAAAGTAAAGTAAAGTAAAGTAA 2351
515 uLysGlnGluLysLeuGlnAlaLysTyrValAsnArgGluLeuValLeu 532
2352 AAAAGAACAAAATTTAGTAGCTAAATATGAATAAAGATTAAGTAAAG 2401
532 snAlaAlaGlnGlnGluAlaGlnLysAsnTyrThrValValLeuThrAla 548
2402 ATGCACATGCGCCAGAGAGTGTAAATATACTGTGAATTAACAGCTTAA 2451
549 SerGlyGluLysGluAlaLysAlaThrLeuAlaLeuGlnLeuLysAla 565
2452 TCTGTAAAAAAGAGTGAATCTTAACACTTACTTAAATTAAGTCTCC 2501
565 oGlyLysPheSerLysPheGluValArgGlyLeuAspThrGluLeuAsp 582
2502 AAGCTATATCTTAATTCGATTCGTGTGTTAGAAAATGAATTAAGATA 2551
582 yTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerValLeu 598
2552 AATACGTTACTGAAGAAACAAAAGAACAAATGCGTAGTTCTGTCTT 2601
599 ProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaLeuLeu 615
2602 CCAATGATGCTAATGTTAGTTTAAAGAGAAAGAACGACGCAACACT 2651
615 uLysValThrThrThrAsnLysGlnGlyLysGluValAspAlaThrAsp 632
2652 TAAAGTACTACTACAGATTAAGATGTAAAGTAGATGACAACTTCAAG 2701
632 laGlnValThrValGlnAsnAsnSer...ValIleThrValGlnGlnGly 647
2702 GTCAAGTACTGTAAATGATGCTGCTGCGACAAATTACTGTAGAAATGA 2751

```

```

648 AAlaYsAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLe 664
|||||
2752 GCAAAAGCGGTGAAACATATATAAGTAACTGTAGTAGCAGCGTAAGCT 2801
664 uUleFThrHisSerPheLysValValAspThrIaPofThrAlaLysG 681
|||||
2802 AATCACACTCATTCCTTTAAAGTTGTTTACACAGCACCTGCAGCTAATA 2851
681 lYLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProasn 697
|||||
2852 AATTAGCACTAGATTTCACAGACATCTCTTAACAGAGTGTCTCAAGCT 2901
698 AAlaSPLeuLysAlaAlaLeuLeuasnIleLeuSerValAspGlyValPr 714
|||||
2902 AGCGAATTTAAAACTGCACCTCTAATAATATCCATCATCAGTTGATGCTTC 2951
714 oAlaThrThrAlaLysAlaThrAlaSerasnValGluPheValSerAla 731
|||||
2952 AGCAACACACTGCGGGTGCACACTGTAACGTACAGCGAAATTTGTTCTGCAG 3001
731 sPThrAsnValValAla...GluAsnGlyThrValGluValAlaLysGlyAla 746
|||||
3002 ATACAAATGTTGTAGTGAAGAAACGCTTAAGTTGTCACAAAAGTTCT 3051
747 ThrSerLeuTyrValLysAsnLeuThrValValLysAspGlyLysGluG 763
|||||
3052 ACTTCATCTTTGTTGTAAGAACACTTACTGTATAAAGAAGTGAGCAACACA 3101
763 nLysValGluPheAspLysAlaValGluValAlaValSerLielysGluVal 780
|||||
3102 AAAAGTTGAGCTTGACACAGCAGTACAGTACAGTGTATTATAAGAAG 3151
780 lAllys 781
:::|
3152 TAAAA 3156

seq_name: gb_ba:BFH249446

seq_documentation_block:
LOCUS      BFH249446      3159 bp      DNA      BCF      01-JUN-2001
DEFINITION Bacillus thuringiensis slpa gene for surface-layer protein.
ACCESSION  AJ249446
VERSION    AJ249446.1 GI:6580204
KEYWORDS   slpa gene; surface-layer protein.
SOURCE     Bacillus thuringiensis.
ORGANISM   Bacillus thuringiensis
            Bacteria: Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE  1 (bases 1 to 3159)
AUTHORS    Mesnage,S., Haustant,M. and Pouet,A.
TITLE      A general strategy for identification of S-layer genes in the
            Bacillus cereus group: molecular characterization of such a gene in
            Bacillus thuringiensis subsp. galleriae NRRL 4045
JOURNAL    Microbiology 147 (Pt 5), 1343-1351 (2001)
MEDLINE    21220765
REFERENCE  2 (bases 1 to 3159)
AUTHORS    Mesnage,S.
TITLE      Direct Submission
JOURNAL    Submitted (14-SEP-1999) Mesnage S., Bacteriologie / Mycologie,
            Institut Pasteur, 28 rue du Dr Roux, PARIS 75724 Cedex 15, FRANCE
FEATURES   source
            1..3159
            /organism="Bacillus thuringiensis"
            /strain="NRRL 4045"
            /sub_species="galleriae"
            /db_xref="taxon:1428"
            270..2751
            /gene="slpa"
            270..275
            /gene="slpa"
            /note="contains 3 SLH (S-Layer Homology) motifs towards
            its amino-terminus"
            /evidence=experimental

```

seq	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465
-----	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----


```

          SSKYIAVAKGAVVGKGGCKNPEYDEKGIDRASPSPAMVGVYNLKEKYDGLTYRKFDQD
          RGHNGEKANILIVLVGIISVTGCGKMEKNPESRAEALAEFIALDPDKTAKRENSDAKYT
          NVATPERPOLTLTGTLGNLTLEADVTLLEGKMLALEOKSAVVLYLSKRIAPNKEL
          PVKAGKNTFTFYKVYEVVKRLAREQLTEITDDDRADAQVAVKINDEKGNADIETLDIAGHD
          VKFEANNIGDTGPANIIEFGVGAESTETGTGLADGAEFLKNEVOYTKRGDIYVSNGLIE
          VKNLEADPATIKEDVAFVDNDKAGVNAAKPLSGDPTLNSTKATIVAGERKAGIHKKVAOI
          NKRKKADVPSAISLKSSNPAGIVSYKNEKTAEEAASGDFLHVXGDVYRTGEDVEVKTOT
          RKLTITTVKANPDOLKRVNGKELPVTFHTVTDQYGDFEGNSAIKEVPFGOVVYLDVT
          TTNGDSIGTSISKVGENNGAGTHFNPNPASGSYSILHYEUTKSNGHEARPLELY
          SKAOKEAADDTLLGAGNTAYATOLSNTTREBVIADADKLGERFERVNDKIATAKIEG
          KTLKVTYKTAGSDVDNLGINKLINSTEHNHATVIVETOINOITVYKERDEVGEENKVINI
          DRYLDVETKAGDVINDGINKLINSTEHNHATVIVETOINOITVYKERDEVGEENKVINI
          YVIVAVKRNVSKEGNDLFKFLEFDEGTATNRKNDVFKEGYTTFABDKGITVIFKVMKDVRAP
          TTEYEKANDVINVYKEEI"
BASE COUNT      1151 a       501 c       670 g       945 t
ORIGIN

alignment_scores:
    Quality:      760.00           Length:      839
    Ratio:        1.532            Gaps:        34
Percent Similarity: 59.118     Percent Identity: 31.943

alignment_block:
US-09-754-947-1 x BU038842 ..
Align seg 1/1 to: BU038842 from: 1 to: 3267

1 AlagilylsthPherproaspvalProAlaasphistrglylleaspse 17
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
441 GGAGGTAATTCATTCGCCAGACGTCGCAGCTGGCAATTCGGCAGAAATATTC 490
17 rtleasnryleuValglulsglylaValalylsglylsGlyasnAsplyslgym 34
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
491 TATTACTACTAGTAGATGATAAGGTGCATTCGATGAGTAAAGCCAGACGGTA 540
34 eRphecluprocllylsgluleuthArgAlaGluaLaalathrmctmet 50
::: ||| :::::::::: |::::|::::|::::|::::|::::|::::|
541 CATATGCTCACCGCATCGATCGATCGATCGATCGATCGATCGATCGATTTTC 590
51 AlaGlnlleuasnleuprroileaspIysaspAlalySproserpheal 67
:::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
591 ACATAAATTTTAATTTACCAGTGTGATGAAATGCCCAACCTCTTTCAA 640
67 aaspSerGinglyGlntPrTyThrProBehlealaValaValglulya 84
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
641 AGATGCTAAAATCTTTGGCTCTCAAAAATATATTTGGCAGCATGTGAAAAAG 690
84 laglyValilleysglyThngly..AsnGlpPhecluproasn glylys 99
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
691 CTGGTCTCGTTAAAGGTGATGGAAAAGATTAATTTCTATCTCTGAAGAAGAG 740
100 lleaspArgValserMetAlaserLeuleuValglualatyrllysleus 116
741 ATTGACCGAGCTCTCTTCGCGCTATGAGTAGTAGTGAGCTTATTAACCTAAA 790
116 pthrlyValasn glyThrProAlathrlsyPheIysaspReugluThrl 133
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
791 AGAAAATGATAGATGGCACAATTAGTTCAAAATTTTGATGATTTAAAGAGA. 839
133 euasnrPrpGlylsGlylsAlaasnIleueValJluLeuglylieser 149
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
840 .CATTGGGGGTGAAGAAAGCTAACATCTCACTTAACCTTGAACCTCTG 887
150 ValGlyThrGlyaspclntPrpGlyupPolyslysthyValThrltyslaet 166
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
888 GTGAGTACGTGGTGGTAATAGGAGCCGAATAAATCTGATCTCGTCGACAG 937
166 ualaIaaginPheIIeAlalyshrsPlysglnPheglyThrGlualaA 183
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
938 AGCAGCTCAATTTATCGCGTTAACATATAAAATAT.....G 915
183 lalysValgluSerAlaValaValThrGlnlylsValgluVallys 199
```


[illegible]

	Ratio: 1.497	Gaps: 41
Percent Similarity: 56.818	Percent Identity: 31.477	
alignment_block:		
US-09-754-947-1 x D86346 ..		
Align seg 1/1	to: D86346	from: 1 to: 2842
1	AlaGlyIysThrPheProAspValProAlaSpHisIstPcLyIleAspSe	17
264	GCAGGAAATCATTCCTCCAGACGTCACACCTGGTCTTGGAGTGCAGAGTA	313
17	IleAsnIyrIleuValIGIuysGValIaValIySGLyAsnAspIysIym	34
314	CATTATTAATTATTAGTACGCAAAAAAGCAATCGAAGGTAAACCGACGGA	363
34	ePheGIuProGlyLySGIuIeuThrIrfAlaGluAlaIaIaIaIaIaIaMet	50
364	CATTGCGTCACACGTAACACATTCACGTCGCTTGTGACGGAATAATTATG	413
51	AlaGluIleuAsnIeuProIleAspIysAspAlaIysProSerPheAl	67
414	GCAATTAATTAGCGCTTGAAGTTAAAGAGGCGTCACAAACATCTTTCAA	463
67	AspSerGInGlyInIrrPyrThrProPheIleAlaIaIaValIGIuIysA	84
464	ACATCTCTACAGATTTCATGCGCTGCTCAATTAATATTCGACGCTGTGAATAAG	513
84	IaGlyValIleIySGlyIthrGlyAsnGly... PheGIuProAsnGlyIys	99
514	CTGGTGCTATTTCACAGCGGATGAACCTGGTAGTTTAATCCAAACACCAA	563
100	IleAspIyrValSerMetAlaSerIleuIeuValIGluAlaIyrIysIleAs	116
564	ATTATATCGCTTCATATGCGCAACATGATGCTTAACACATACAAATTAGA	613
116	PthrIyrValAsnGlyIthrProAlaIthrIysPheAlaIysAspIeuIuIhrL	133
614	TGGTAAAGTACCGCAATTAGAACTAAATTTCTTGATTATAAAGAT..	662
133	euaSnTrpGlyLySGIuIySAlaSnIleuValIGIuIeuGlyIleSer	149
663	.. CACTGGGGTGAAAGACCTCAACATCTTAGTTCGTTTAGAATATCACT	710
150	ValGIythrGlyAspGlnTrpGluProIySlyThrValIthrIySAlaG1	156
711	AATGCTACTGCTAAATGCTTGGGACCGACAGATAAATCTGTAATCTGTCAGA	760
166	uAlaIaGlnPheIleAlaIySlyThrAspIySGlnPheGlyIthrGluAla	182
761	ACCTCTCAATTCAATTCGTAAGACTGACATGCCATCTGGACCAAAAAACAG	810
183	.. AlaIySValIGIuSerAlaIySAlaValIthrTrpGlnIySValIGluVal	198
811	AAGCTTAAGCTTCACTTATTAAAGCAATCAATGCTTAAGAATCGAAGTT	860
199	IySPheserIySAlaValIGIuIySleuThrLySGIuAspIleYsValIth	215
861	AAACTGGGTGCTACAGTAACAGAAACCTGTGAAGAACGGAATAATTAC	910
215	IAsnIySAlaSnAspIySValIeuValIySGluValIthrIeuSerG	232
911	TCTTTCTCAAGAGTGTATTAACGTAAACAGTGTGAACCAAAAATTCGGTG	960
232	IuAspIyAlrSerIaIa.... ThrValIGIuIeuIyrIserAsnIleuAla	246
961	AAGCCAAAAAATCTGTATCTTTAACTTTAGAGATGCTAAAGACTTGAG	1010
247	AlaIySGlnIthrIyThrValAspValAsnIySValIGlyIySthrGluVal	263
1011	AAAAACATGCTTATGCTTGAACGTTCAAAAA.....	1044
263	IAlaValIGlySerIeuGluAlaIySlyThrIleGluMetAlaAspClnIthrY	280

```

1044 .....TTAAACACCTGCTGAAGCTAAAGAAATTCACAAAGCATTAAGAGTTA 1089
280 aValaIaIaSpGIuPProThrAlaLeuGlnPheThrValLysAspGIuAsn 296
1090 TCCTCTTTAAAGTGAAGTAGCT.....CCAACT 1118
297 GLYThrGIuValValSerProGIuGIleGIuPheValThrProAlaI 313
1119 GTATCTACAGTATCTACTCCAGACGA...AAGTTAAAGTTGTAATTAG 1165
313 aGIuLysIleAsnAlaLysGIuLeuThrLeuAlaLysGLYThrSerI 330
1166 TGAATAATTAGATGACGCAAAACTGTAAT..... 1196
330 hrThrValLysAlaValTYrLysAspGIuLysValAlaGIuSer 346
1197 .....GTAATTGTTAACGGC 1211
347 LysGIuValLysValSerAlaGIuGIuAlaValAlaSerIleSerAs 363
1212 CAAGAGTAGAGTAAAGTTGAAATAATACCTTCACT.....TCATA 1255
363 nTPPrThrValAlaGIuLysAsnLysAlaAspPheThr..... 375
1256 CCAAGCTTTAAACCTTCAAAATGTTAAAGATTCATCATCATGTAAGT 1305
376 ..SerLysAspPheLysGIuAsnAsn...LysValTYrGIuGIuLysAsn 390
1306 GTGCTACAGATTAGCTGATACACTATGCAAAATGTACGAAGGCAAGCT 1355
391 AlaTYrValGIuValGIuLeuLysAspGlnPheAsnAlaValThrThrGI 407
1356 ACTTAT..... 1361
407 LysValAlaIuTYrGIuSerLeuAsnThrGIuValAlaValLysAsp 424
1362 .AAGTGAAGAAAGATACTACAGCTCCAGAGTTAAAGATGTA..... 1403
424 lathrGIuLysValThrValLeuSerAla...GIuLysAlaProValLys 439
1404 .....AAGTTAAAGAAATTAAGTCTGTAAGGCAAGCAACATTAGAG 1445
440 ValThrValLysAsp..... 444
1446 GTTACTTCTCGAAGAAGCTGTAACAATGTCGAAGCAACAGTTGTAAGT 1495
445 ..SerLysGIuLysAlaLeuValSerHisThrValGIuLeuGIuAlaPheA 461
1496 TAAAAAAGCTAAAGAAATTAATTAATCTACAAATGCTGTGAC.....G 1539
461 lalGIuLysAlaMetLysAspIleLysLeuGIuLysThrAsnValAlaLeu 477
1540 CTCAGATTAAGCAAGAGCAGT...ATTGAAGTAAACAGCGCTCTTAA 1586
478 SerThrLysAspValThrAspLeuLysVal..... 487
1587 GAAAAAAGAACTGTACAGATCTACAGTTGAATTTGTTGCTACAAAGA 1636
488 .....LysAlaProValLeuAspGIuTYrGIuL 497
1637 TGCACCTAACACAGTAGCTTAAAGTATCTAAAGCAGTAAAGTTTCAA 1686
497 LysGIuPheThrAlaProValThrValLysVal..... 507
1687 AAGAGCTTGTACTCTCTAATCTGTAAGTGAAGCTGATCAAAATATA 1736
508 .....LeuAspLysAsp.....GIuLysGI 514
1737 GCAGGCACATTCACATTGTGCAAAAGAACTGACAAACGGAATA... 1784
514 uleuLysGIuGIuLysLeuGIu...AlaLysTYrValAsnArGIuLeuV 530

```

```

1785 .CTTCTGTGATTAACTTAGATACAGTAAAGATGTAACTGTGATGTAA 1833
530 aLeuAsnAlaIaGIuGIuAlaGIuAlaGIuAsnTYrThrValValLeuThr 546
1834 CTGTTCCACACAGCAAAAGATTAACAAAAAGTTTCAATTAAATTCACCA 1883
547 AlaLysSerGIuLysGIuAlaLysAlaThrLeuAlaLeuGIuLeuLys 563
1884 GCAAAAGGTACTACAAAGAGCT...GCTACAAAGGTTCTTAAAGA 1930
563 salaProGIu.....AlaPheSerLysPheGIuValArGIuL 576
1931 TGCAGCTGTGAATAGCTGTGATGATTAAGTAA...GAAGTAAAGTTG 1977
576 euAspThrGIuLeuAspLysTYrValThrGIuGIuAsnGIuLysAsnAla 592
1978 TAGAAAAAGAGCTGAAAAA.....GAACAGATTAAGTAGCT 2015
593 MetThrValSerValLeuProValAsp...AlaAsnGIuLeuVal..... 606
2016 CCAAGTAAAGAGCTGATGCTGATGAATGAAGCGCAAAATTAAGTTACAGT 2065
607 .....LeuLysGIuAlaGIuAlaIaGIuLeuLysValT 618
2066 AAACCTTGATTAAGAGATTAAGAGTGCTCAGGTTCGATTCAGCTGCTA 2115
618 hrThrThrAsn.....LysGIuGIuLysGIu 626
2116 ATGTAATAATTAATTAACAGTTAGCTGTGCTTAATTAACCAAGAGTACTTA 2165
627 ValAspAlaThrAspAlaGIuValThrValGIuAsnAsnSerValIleTh 643
2166 ATTGTATTAGACGGGTACAAAGCAATTATTGAATTACATTACATTACATT 2215
643 rValGIuGIuGIuLysAlaLysAlaGIuThrTYrLysValThrValVal 659
2216 TTTC.....GAAAAATCAGAAACAGTAAATTTACAGTAGCTA 2253
660 .....LeuAspGIuLysLeuIleThrThrHisSer.....Phe 670
2254 ATGTTCGAACAAAGATGAGTAAAGATTAATTAATTAATTAATTAATTA 2303
671 LysValValAspThr...AlaProThrAlaLysGIuLeuAlaValGIuP 686
2304 GAAGTTAAAGTATCAAAAGCTCTGAATTTAA.....TCAGCTAAAT 2347
686 eThrSerThrSerLeuLysGIuVal.....AlaProA 697
2348 TACAAAGTGTGATGCTAAAGAAATTAATTAATTAATTAATTAATTAAT 2397
697 snAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGIuVal 713
2398 ACGTTGATACACAGTCTACAGTTGATTTA.....AAGCGTGA 2438
714 ProAlaThrThrAlaLysAlaThrAlaSerAsn..... 724
2439 GAATTAAGCTGTGCAAAAGCTGACCAACAGCAGAGCTGCTAAAGAGCT 2488
725 ..ValGIuPheValSerAlaAspThrAsnValValAlaGIuAsnGIuTYrV 741
2489 AGTTCTTAAGTAAAGAGCTCCAAAGATGTTACTTAGCAACTGGTACAG 2538
741 aGIuLysAlaLysGIuAlaThrSerIleTYrValLysAsnLeuThrValVal 757
2539 TTACAGTAAAA.....GCTAAAGAACTAGAA..... 2564
758 LysAspGIuLysGIuGIuLysValGIuPhe.....AspLysAlaVal 771
2565 .....AATATGCTAAATATATGATTCAAAACTACTGACAAAGCAGA 2605
771 lGIuValAlaValSerIleLysGIuAlaLysProAlaThr 784
2606 AAACGCACTTGTGCTTTC.....AAACAGTAA 2636

```


ORGANISM	Plasmid Bacillus anthracis
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
AUTHORS	1 (bases 1 to 96331) Okimaka,R.T., Cloud,K., Hampton,O., Hill,K.K., Keim,P., Lamke,G., Kumano,S., Mantler,D., Martinez,Y., Svensson,R., Tatum,L.R., Brown,A.E. and Jackson,P.J.
TITLE	Direct Submission
JOURNAL	Submitted (01-NOV-1999) Bioscience Division, Los Alamos National Laboratory, M888, Los Alamos, NM 87545, USA
FEATURES	Location/Qualifiers
source	1. 96231
CDS	<pre>/organism="Bacillus anthracis" /plasmid="px02" /db_xref="taxon:1392" complement(19. .477) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-01" /protein_id="AA013606.1" /db_xref="GI:6470152" /translation="MREGGDDVSLKKEVKDLKQSYNTMKQOMVLKLNMLIAQ SGSNLSLEKLEQEVKSEIASVTKSEKGFPTVREPIEENVOEHQRIQREVKTTSF VKKPLMTDDVOIYMKKGKPHQVPTRLGTGQETEMEQIPEHRRKELE" complement(482. .1018) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-02" /protein_id="AA013607.1" /db_xref="GI:6470153" /translation="MGKSPSIIILTSQYTIIPNKSFTTYVDYCMROKALLKDGHTLNT ELTELSRLQNAIDSEPMRSGPSNOKKVSSEKEAKVLLTSPNSFMTTDFEKYV SYNGRKYALESKKMTKMTTEEEVSLNKHIEDLSPMEKKGKGIKABEILISGVSK DKNVITSTDLNFKDRIK" complement(1037. .1633) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-03" /protein_id="AA013608.1" /db_xref="GI:6470154" /translation="MTYVKLFQFRGNQIHVDIVDFORQYPEYPREADVLRIHAIVLLD EKMEREGGDDVSLKKEVKDLKQSYNTMKQOMVLKLNMLIAQSGSNLSLEKLEQ EVKSEIASVTKSEKGFPTVREPIEENVOEHQRIQREVKTTSFVKKPLMTDDV QIYMKKGKPHQVPTRLGTGQETEMEQIPEHRRKELE" complement(1645. .1932) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-04" /protein_id="AA013609.1" /db_xref="GI:6470155" /translation="MEIHVRNVDPYHLKEIDRCKRIGTKKGLGRYRYRWETINMFEQH FNOESIRNKEDKDEAVSNVSTILDKQSKRLQETIDATHELVAAMTKLME" complement(2240. .3088) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-05" /protein_id="AA013610.1" /db_xref="GI:6470156" /translation="MSIKFWMNKGRRKRPANTKRAVLTFLGFSVTMFVVLCSIPVFSG ATYKFEEMKIGEQSLSTYIAVAREKYNPDQTLIDYELRADNSOILSNMKYK ENKYIKQDNVNTKTYIRASDNTIYVISENVPEEFGVSVYKPEYIHPELODVDL KRSKMYVLENKLIINRLKRSKDPYREVIAPSOALRKIEKKMKDSAMQOL KIKNEQJLTEMETQTEGEKVKTKNTINSNESTINNHQEKIDVLKEDIKKKERTQLD EKRTI" complement(3085. .3360) /codon_start=1 /transl_table=11 /evidence=not_experimental</pre>
CDS	<pre>/product="px02-06" /protein_id="AA013611.1" /db_xref="GI:6470157" /translation="MLSELKPNHDYAKGKYLILSRKKGVKROKFIETPTWEDY NGEKEVEMIVREIOPVNGKEKRYNCKLENIHAQVSVYVNGERNK" complement(3449. .4057) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-07" /protein_id="AA013612.1" /db_xref="GI:6470158" /translation="MNNLSKITSGALVLSLSPNGIMAYNTTEPDKKDKRYVYSEK EKQIDGLKVKLEQKDKKINEGAGENKSTLIDQNTYREVAQFVAIYDYSVN KGERNNLKITDKKVVLDIAPNTDLDGPNKSHVKNKAIYINSGDVSKKCTALD IETTIEGLENKQTTINSYVKITLERQGEIEIKVEYVNPYVKR" complement(4083. .5225) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-08" /protein_id="AA013613.1" /db_xref="GI:6470159" /translation="MAAQETAIIDKYKKYKRIKWIVRLIGSGSVIAAITLLIIVS MAIRGQSSTGTPNGISGTRATYVKNLPPRYVKRQANVEQCAAGVPELVPYLAIIIM VESNGISEKLPDIMOSSESQGNAMNTISPKDSIYGVNHLKAGDDADAMVLINDLLA IVQTYNFGRNIVHMLAANNKTHSIQIADYSLIYVAPAGNNGNTTIGYSQVAAVYN GGYRIYNGNGFFYAEWKQYLSFDGAGTSGQIPGSEFEKVMQDEVLYKNGNPYVMG GKSSQGFDCSGLTGYWAKYTAGITIPISATQYDFTVEYDPKADQPGDLVFFRGYGG PMHVSHVGIYDIAITMYDNSGSGVGHQPTSSYMOHVGDIRVVR" complement(5227. .7158) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-09" /protein_id="AA013614.1" /db_xref="GI:6470160" /translation="MKLGKKEQOKYRTBEGYDIDFTISRVPQSGIKFENEYIAGDCYV ACLIHVSLAEDIPPLMTLTALMNNKDTISSVDVATNAKEEVPVDINKSTIELKDRMTSE RSTDPBDAHMELOMTDFARSITQOGEIYLVKRIYIYDVLQLEKRIDIKKEI AGONYKAAQYVFFKOKEMOTLPASVDQLEIYGVSGYLPKSNIGFCGIPFHODLKD PRGIYIGORTSGTGAFILDPFSGTRGTSRSPGFGMGKGLSTLQKLTNTEGAVDCE TSGFDKARYTYVVOQGGKIIDLSGVDSQSEVENMTNPLPEFATKLTNTEGAVDCE SETOHISKVTYNNIRLNFTEFDTVIOEFKHLAFITENGLYPOKSDRPYVYTGTP NMYPLIEDDYKYLINLKLAGATPOHNRDLKILQVEDMTIYVGMFNGHTLNFE NQOIVYFEDYKYSYKQSVNQLFTALTLINSHALKRQDKYLRBEKNLSIEDVKY FMVFLDECHNVINSQNEFANRYVLEFEREMKFFAGIIFAPQSPOMLPPGSTVDVS TMKAIPELTQYKIFLNMNSVLETLKSVLGESELTSEFRILPELKGCAIVQVSTET YVMEPDDEKQLEPRKGGQ" complement(7178. .7846) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-10" /protein_id="AA013615.1" /db_xref="GI:6470161" /translation="MKVATAEAKKYNRKQIASEKSNQKOSTISTIPASFKNVDASIV DIAPYKRLBOKTGILVDHRRNLOVYLKVTTLILSNQODLKRPMQNLSTLCRYHEP FKILSLTSTETTEQOVYWRKRALRFGQMSQSEVSKKEEELMYORYSLAINTLNKLV WVENKLKELAFIIVYGNMETELIKNVKDKMRYGQHPLQNMKKAKEVERKLLFKLQNM NSEM" complement(7910. .8224) /codon_start=1 /transl_table=11 /evidence=not_experimental /product="px02-11" /protein_id="AA013616.1" /db_xref="GI:6470162" /translation="MEMMRNPKNTKOEIKLAFYIYIDGAILMLMVLVSLYVAVG GFGRIIMFYLVFGTFGLICIKPHNSPTNNIFVIIDMLMDKRNKHPIDVNTISSETK RK" complement(8227. .8430) /codon_start=1</pre>

```

/transl_table=11
/evidence-not_experimental
/product="px02-13"
/protein_id="AAFI3617.1"
/db_xref="GI:6470163"
/translation="MNMWIFAILMLGVAIVLSIATPEFKOITTFKRRKKKEPWVIT
TNAEYLDREDFWKCEKQDKKEDH"
/complement(8432..8686)
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="px02-13"
/protein_id="AAFI3618.1"
/db_xref="GI:6470164"
/translation="MKVIDIANRRRIYEFEMKQOELRASILMTVAGFIIAFAILVFOIS
FELGHLYHYIVTFAFVLTSLHLSSNNKLARKIEKKOQY"
/complement(8704..11562)
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="px02-14"
/protein_id="AAFI3619.1"
/db_xref="GI:6470165"
/translation="MATMKIKEMFROSKKALVIGLAVITLVSEFTMPPIAYAFDWE
FLGLKDVPTKEVAFLOENWTOYADPLGINSIOGGLIKLLYIVLSALEGLIPDFS
FEDILKADGNDFASSILKLEFLVLTIVLWLGRTIIQHKPPPKSVGNVILYMG
LLGLENLMADQKMSDTSFSETNTSKAKDGLADLVKONTADILYLSKTEFEPTQS
KRENPASISTDSKPKNGMSKIDFLKADGIVPTKVIEMLTADIKPEYELVYKI
TNDKPEYIEKINDSPNPKFPGGVVRYPMNGTLEFAGLAVLYLFTVPEVMA
TTFETAMKKIAPIVFMDIETGEKTKVYIDIFKGFVFAFTGLNLRVYTIIVYLA
EKDINAPLIVAMVCLITVALINGSESLIKRGVDYGLAKGHLGMVGGAAMAKGAI
KAGMFSGAKNLMGSGSQTREDTPKRSSEMSNHESAGEOKNTAKGSDVSKTGS
ANTGAAMSYAANRGVGGMNDAGNKVAVETVGCAYDKCEKESGVAKGSVDKCTVS
ERKGOAECKOAEADROTLNTIGKGIIONNANGIEPSAPKIANRPVNASISER
OOGNIGKEROTAREIEIOLNKLNTGKEGQGNVSEKQTEREIEIOLNKLNTGST
EGOONVNGEROTAREIEIOLNKLNTGKEGQGNVSEKQTEREIEIOLNKLNTGST
GCTGINSOKKAVATRODVONNNRGSEAVDSNVRASTTVORDVONNONGNVRGSEA
VESNVRASTTVORDVONNNGSFAVNHVTRAGATTVORDVONNNGSSEVADVAVG
TTTVORDVONNNGSEVRSRLRPKRGSTONRFDQLRR"
/complement(11588..12379)
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="px02-15"
/protein_id="AAFI3620.1"
/db_xref="GI:6470166"
/translation="MLFNVLQSYQIEKLYDKATEIKDNCONHLYIMSTNKDTIEE"

CDS
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="px02-15"
/protein_id="AAFI3620.1"
/db_xref="GI:6470166"
/translation="MLFNVLQSYQIEKLYDKATEIKDNCONHLYIMSTNKDTIEE"

alignment_scores:
Quality: 487.50 Length: 564
Ratio: 1.573 Gaps: 20
Percent Similarity: 54.965 Percent Identity: 28.369

alignment_block:
US-09-754-947-1 x AFI88935 ..
Align seg 1/1 to: AFI88935 from: 1 to: 96231

3 LysThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsn 19
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38005 AAGACATTACGATGTACCA.....AACTGGCACAACAATCAGATTAA 38048

19 nTyrLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheG 36
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38049 CTACTTAATGAAAAA...GCACCTGATGGAAGAACACAGATGCACTTTT 38095

36 LpProGlyLysGluLeuThrArgAlaGluAlaAlaThrMetMetAlaGln 52
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38096 CACCATCTCAAAAAAATGACCTGCTTCAGTCAAAATTAATGCGCTATG 38145

53 IleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSe 69
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

38146 GTACTCGGTTTACAGATTACAAACAGCAAAACCGTCATTCCAGATGC 38195

69 rGlnGlyAsnTrpTyrThrProPheIleAlaValGluLysAlaGly 86
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38196 AAAAAATCATTTGGCATCTCCATATATTCACAGCAGCAAGGCGTGG 38245

86 aIleLeuGlyThrGlyAsnGly...PheGluProAsnGlyLysIleAsp 101
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38246 TAATATATGAGATGGTAGTGGTACTTAACTTAACCCATCAAGGACATTGAC 38295

102 ArgValSerMetAlaSerLeuLeuValGluAlaTyrLysLeuAspThr 118
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38296 CGTGCTCTATGCGATCTATGCTTGTGAAGCTTATAGCTTAATATATAG 38345

118 sValAsnGlyThrProAlaThrLysPheLysAspLeuGluThrLeuAsn 135
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38346 AATTAATCGGATTTACCAACGCGAGTTTGAAAGCTTAAAGG...CAT 38392

135 rPcGlyLysGluLysAlaAsnIleLeuValGluLeuGlyIleSerValGly 151
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38393 GGGGTGCAAAACAGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 38442

152 ThrGlyAspGlnTrpGluProLysLysThrValThrLysAlaGluAla 168
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38443 ACTGCTACCGATGGAAACCAATGAAATGTACACGTCGACAGACGT 38492

168 aGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluAlaLysVal 185
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38493 ACAATTAATGCGGCAACAGACATGAAA.....AAG 38524

185 aGluSerAlaLysAlaValThrThrGlnLysValGluValLysPheSer 201
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38525 CAGACACTTCAAAAAAGATGATGATGAACA..... 38556

202 LysAlaValGluLysLeuThrLysGluAspIleLysValThrAsnLys 218
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38557 .....CAFTTCATTACATACCATCAGCCATCCATCATCAGGGGT 38597

218 aAsnAsnAspLysValLeuValLysGluValThrLeuSerGluAspLys 235
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38598 TACCTCAATCAATCAATGCGCCCAATCATCTGTTAAACAAACACCGG 38647

235 rGSerAlaThrValGluLeuTyrSerAsnLeuAlaLysGluThrTyr 251
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38648 CAGATGCTGTGATATAAATTGTACGAATATAGGATATAATGCACTCA 38697

252 ThrValAspValAsnLysValGlyLysThrGluValAlaValGlySer 268
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38698 TTATATGAA.....AAAGAGA...ACCATTCACAGTAC 38729

268 uGluAlaLysThrIleGluMetAlaAspGlnThrValAlaLysGluP 285
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38730 GTTACTACATATCCAGAACGCTCACATTCATCTAAATATTAAGTACGC 38779

285 rOThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValAl 301
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38780 ATTCCTCCCAACAGTAACAGTATATAGAGAAAAGAGAGT..... 38820

302 SerProGluGlyIleGluPheValThrProAlaAlaGluLysIleAsn 318
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38821 .....TGATTCGCGATCTGTCAATATGCTAGTTCATGCTGTGGA 38861

318 aLysGlyGluIleThrLeuAlaLys..... 326
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38862 TAAAAACCAATTAACCTACCAAGAAACAAATAACTTCTTAAGAGCA 38911

327 .....GlyThrSerThrThrValLys 333
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38912 AAACATTTATTCATCCAGTCAATGCTGATGATAGATGCTGAGACATAA 38961

334 AlaValTyrLysLysAspGlyLysValValAlaGluSerLysGluVal 350
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
38962 GGAATTTATATATGAAGAACCCAGTTGTTATACGACACA...GCA 39008

```

350 sva|Ser|a|Glu|Gly|Ala|Ala|Val|Ala|Ser|Leu|Ser|Asn|Trp|Thr|Val|Ala 367
:::|
39009 CGTA..... 39012
367 lae|Glu|Gln|Asn|Lys|Ala|Asp|Phe|Thr|Ser|Lys|Asp|Phe|Lys|Gln|Asn|Asn 383
:::
39013 CAA 39015
384 Lys|Val|Trp|Glu|Gly|Asp|Asn|Ala|Trp|Val|Gln|Val|Glu|Leu|Lys|Asp|Glu 400
|||:::|:::|:::|:::|:::|:::|
39016 AAATTGTTGGCGAAAGACTCCATTACTGCTCATTTAATTAACACGGGATGC 39065
400 n|Phe|Asn|Ala|Val|Thr|Thr|Gly|Lys|Val|Glu|Trp|Glu|Ser|Leu|Asn|Trp 417
:::|:::|:::|:::|:::|:::|
39066 TTAACA..... AGACCTGGAAAAAATGCACACAGACTCTCTAGGAAAA 39109
417 Lu|Val|Ala|Val|Ala|Asp|Lys|Ala|Thr|Gly|Lys|Val|Thr|Val|Leu|Ser|Ala 433
:::|:::|:::|:::|:::|:::|
39110 GAGTTGAATTTGGCAAAAAAATAAGGCGATATTT..... 39147
434 Gly|Lys|Ala|Pro|Val|Lys|Val|Thr|Val|Lys|Asp|Ser|Lys|Gly|Lys|Ala|Le 450
|||:::|:::|:::|:::|:::|:::|
39148 GTAAGTATTCATGCTTAATGGTTCATAGCTAAGCACA 39185
450 u|Val|Ser|His|Thr|Val|Glu|Leu|Ala|Phe|Ala|Gln|Lys|Ala|Met|Lys|Ala 467
:::|:::|:::|:::|:::|:::|
39186 TGGTACAGAACTTTCTACTACAAAGCACCAACCAAAATCAATCCCGT 39235
467 sp|le..... Lys|Leu|Glu|Lys|Thr|Asn|Val 475
:::|:::|:::|:::|:::|:::|
39236 ATGTAATATGATAGTCGTAATCTACTAGCTGAAATAATTCAAAATGCTCTAATC 39285
476 ..Ala|Leu|Ser|Thr|Lys|Asp..... Val|Thr|Asp|Leu|Lys|Val 487
|||:::|:::|:::|:::|:::|:::|
39286 ACACCTCTCTCAACACGCTGATAGGGGTCTAAGATAGCTAATTTATATGCT 39335
487 L|Lys|Ala|Pro|Val|Leu|Asp|Gln|Trp|Gly|Lys|Glu|Phe|Thr|Ala|Pro|Val|T 504
|:::|:::|:::|:::|:::|:::|
39336 TCCT..... CGTGAATAATACATGCCATCTCG 39361
504 hr|Val|Lys|Val|Leu..... Asp|Lys|Asp|Gly|Lys|Glu|Leu 515
|||:::|:::|:::|:::|:::|:::|
39362 TTTTAACAGACGCTGGCTTTAGATAATAAAGCAGACGGAATAAATTA 39411
516 Lys|Glu|Gln|Lys|Leu|Glu|Ala|Lys|Tyr 524
:::|:::|:::|:::|:::|:::|
39412 GATTCACCAAGATGAGACACACGTCGACGCTGACGACATCTAAT 39453
seq_name: gb_ba:AF242295
seq_documentation_block:
LOCUS AF242295 4000 bp DNA BCT 24-OCT-2000
DEFINITION Bacillus firmus surface layer protein (slpA) gene, complete cds.
ACCESSION AF242295
VERSION AF242295.1 GI:7716572
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE

JOURNAL Submitted (06-MAR-2000) Biochemistry, Mount Sinai Medical Center,
One Gustave Levy Place, New York, NY 10029, USA
FEATURES
source
1..4000
/organism="Bacillus firmus"
/strain="OF4"
/db_xref="taxon:1399"
765..3560
/gene="slpA"
765..3560
/gene="slpA"
/function="role in adaptation to temperature shift and pathogenesis"
/codon_start=1
/transl_table=11
/product="surface layer protein"
/protein_id="AAF68436.1"
/db_xref="GI:7716573"
/translation="MAYDKSYRKEFLASVAAAVVTAAPALAAPADAKSDVSS
HMAIDINLYEKGAIOGYPDTEKPNNSITRAEIAVVALNLGLDVSDEVTDKFSD
VPATWMANPYIAIYDQTEGVIDGXYENGTFRPSNTITRQEMAKVVEAIDLEVGKD
LEPTVSGIMSTDYINILASNGVAGMATTPAPGVEVLAQTAFIHRAEVEERIE
VPGILVPVSSVTAVANATLTLYLSDDTHEVLDLALVALNEATEVTFEINDVEEYV
TWVEEIAVASVNAVNGKOLVYNPQPDVASVLTPASSTVKTVGLSINRSTDT
HSNDVYLGAERAVLSDDKTLTVTAPTGQFPGKNDYTVKAKSGQNTIPSYITIS
VDDTAPLVNSVYVVPADTFEVTJSEPIDSLTGLVRLNGQPVASGDALGPNKL
TFARSSVATGNNATYIAGFSDAAGNEFTPTSTTVPVVQDTTALVASLEQVTQKV
RLTFPKNELSGATLSESGTVVTRPNSTGTSNTPVQNLVDWMTGKTYDITLSDAT
YTNNGSEVRFGLIDAFDTVDYNNKNDLYSKSITLTKDTPVAPTVSAALASRQALEV
TISEGVITINPAQVKLRKDAEOTGLSVAKGTQNNVLYVSYSGAEIAGSXYQBLE
ACAVALDNGNANNAVNPBSVSATPAPLANAVANSQTNVTVTAPTGQTTTASLN
HNNEFIDGQAVSSNDITLNTSDITTVSLPESDSVKISGNLFTTNGIALESGALA
TATATVTVDTNTPTLTGAOLVASNAVIKTFPNDLIDALTDAADLIDIDLSNCTVA
YNGCAGSAGTSGDGVVSSGVDSKSLVTVTVSPMSDSNMAIVNNSOISVTTLDTADTILK
DANGVYIRNVNVLVK"
BASE COUNT 1241 a 742 c 788 g 1229 t
ORIGIN
alignment_scores:
Quality: 449.50 Length: 958
Ratio: 0.988 Gaps: 37
Percent Similarity: 47.495 Percent Identity: 22.860
alignment_block:
US-09-754-947-1 x AF242295 ..
Align seg 1/1 to: AF242295 from: 1 to: 4000
1 Alac|Lys|Thr|Phe|Pro|Asp|Val|Pro|Ala|Asp|His|Trp|Gly|Ile|asp|se 17
||| |:::|:::|:::|:::|:::|:::|
864 GCTGATGCAAGTCTCTCGATGTTCTTCATCCACACGCGCTATCGATGATGA 913
17 r|L|Asn|Trp|Leu|Val|Glu|Lys|Gly|Ala|Val|Lys|Gly|Asn|P|Lys|Gly|L 34
|||:::|:::|:::|:::|:::|:::|
914 CATCAACTACTAGTAGAGAAAGGTGCTATCCAGAGCTTACCACAGATGGTA 963
34 et|Phe|Glu|Pro|Gly|Lys|Glu|Leu|Thr|Arg|Ala|Glu|Ala|Ala|Thr|Met|et 50
|||:::|:::|:::|:::|:::|:::|
964 CTTTCAGCCCTAACACACTCTACTCTGTCGACAGAGTGCAGGTGTTCTT 1013
1014 GCTAACACTCTTGGACCTGATGATCTCAATACCACTGACCAATAT 1063
66 e|Ala|Asp|Ser|Gln|Gly...Gln|Trp|Trp|Thr|Pro|Phe|Ile|Ala|Val|Al 81
|||:::|:::|:::|:::|:::|:::|
1064 CAGTGAATGTTCCAGCTACACACTGCGCTAACCCATACCTCCCTCAATCG 1113
82Glu|Lys|Ala|Gly|Val|Ile|Lys|Gly|Thr|Gly|Asn|Gly...Phe|Glu 95
:::|:::|:::|:::|:::|:::|
1114 TTGATCAACATGAGAGGTGATATGATGTGTACGAAGACGATCTTCCGT 1163


```
461 AlaGlnValAlaMetLysAspIleuLysLeuGlyThrAsnValAlaLe 477
||||| : : : : : ||| : : : : : ||| : : : : :
1923 GCTAACCCAGATTAGAAACAATTGCAGCGACAGCGGTGATCCGT 1972
477 userThrLysAspValThrAspLeuLysValLysAlaProValLeuAspG 494
: : : : : : : : : : : : : : : : : : : : :
1973 TCGTAAAGGCTCAAAAGCTACATTCAAAGTA.....ACATTAAAGACC 2016
494 LntYrGlyLysGluPheThrAlaProValThrValLysValLeuAspLys 510
||||| : : : : : ||| : : : : : ||| : : : : :
2017 AATACGTAACAATTCACGTGATGTAATGTAACGTA..... 2051
511 AspGlyLysGluLeuLysGluGlnLysLeuGlnAlaLysTyrValAsnAr 527
: : : : : : : : : : : : : : : : : : : : :
2052 .....ACTTCGACAAAACGTAACAGCTACAGTTTCT.. 2084
527 gGluLeuValLeuAsnAlaLysGlnGluAlaGlyAsnTyrThrVal.. 543
: : : : : : : : : : : : : : : : : : : : :
2085 .....GTATCTTAACCTGGCATTGGTCMAAGTAGACACACTGTAA 2124
544 .....ValLeuThrAlaLysSerGly 550
2125 CTGTGAACGGTGTGCTGAGGTTCTACTACTATCAACAATTAAGGTGCT 2174
551 GluLysGlnAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAl 567
||||| : : : : : ||| : : : : : ||| : : : : :
2175 ACTAAGAGAGATTA.....GTACCTGTAATGATGTTGCTGGGGACC 2218
567 aPheSerLysPheGluValArgGlyLeuAspThr...GluLeuAspLysT 583
: : : : : : : : : : : : : : : : : : : : :
2219 AGTACGTAACCAAAATTAAAGTTCTTGACGAGTGTAATAATCGCAAGA 2268
583 YrValThrGluGlnAsnGlnLysAsnAlaMetThrValSerValLeuPro 599
: : : : : : : : : : : : : : : : : : : : :
2269 GCGCACTGAAAGTCCAGCTTAACAACGATGTTCAATTAAGTTATGCT 2318
600 ValAspAlaAsnGlyLeuValLeuLysGlyAlaGlnAlaGluLeuLys 616
||||| : : : : : ||| : : : : : ||| : : : : :
2319 GTAGATGCTAACGGAATATGTTT...GGCGACATCACTAACGATGTAAAC 2365
616 sValThrThr.....ThrAsnLysGluGlyLysGluValAspAlaThrA 631
: : : : : : : : : : : : : : : : : : : : :
2366 TATTACAAAGTAACGACACTGATACAAATGGTGTAATCGTTAATGCAAGTA 2415
631 sPalaGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGly 647
: : : : : : : : : : : : : : : : : : : : :
2416 AATCTACGCAAAATGCTGATCTGATACGTAATCAGTACAGCAACGCTAGC 2465
648 AlaLysAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLe 664
||||| : : : : : ||| : : : : : ||| : : : : :
2466 AAAAAAGTAGCTAAAGAA...ACTTTACAGTTAAGCTTGGAACTGTAC 2512
664 uIleThrThrIleSerPheLysValValAspThrAlaProThrAlaLysG 681
: : : : : : : : : : : : : : : : : : : : :
2513 TCTAGTACCTAGATGTTGAAGTCATGATCT.....ACATTAAAG 2556
681 LylLeuAlaValAlaGluPheThrSerThrSerLeuLys...GluValAlaPro 696
: : : : : : : : : : : : : : : : : : : : :
2557 CTACTGTGTAACTAAGAAGACCTTATTGACCTGATGCTGCTGAT 2606
697 AsnAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerVal..... 710
||||| : : : : : ||| : : : : : ||| : : : : :
2607 AATGGTGTCTTACGTAAGTACTACTACCTAAGTATGATATCAAGACCA 2656
711 .AspGlyValProAlaThrThrAlaLysAlaThrAlaSerAsnValGlu.. 726
: : : : : : : : : : : : : : : : : : : : :
2657 AAACGGTAATCAATGTTGACTGCTGCAACTCTCTAACACTAATTAAGAA 2706
727 .....PheValSerAlaAspThr 732
: : : : : : : : : : : : : : : : : : : : :
2707 AATTACAAAGCTTTAATCTGTATTAAACGGCAGATCGTAAGCAGCACACA 2756
```

```
733 AsnValAlaLysGlnGlyThrVal...GlyAlaLysGlnAlaThrSer 748
||||| : : : : : ||| : : : : : ||| : : : : :
2757 TCTGTAAATGGTCTGATCTAATGTTGATTAATAAAGATGATGCTAG 2806
748 rIleTyrValLysAsnLeuThrValValLysAspGlyLysGluGlnLysV 765
||||| : : : : : ||| : : : : : ||| : : : : :
2807 CAT.....TCTGATTAGCTGTTAACAAGACGATCACTAATTA 2850
765 aL.....GluPheAspLysAlaValGlnValAlaValSerIle 777
: : : : : : : : : : : : : : : : : : : : :
2851 CACTAGTGTTAATGAAGATCTTAAGATTGCTCAATCGCAATCAGATT 2900
778 LysGlnAlaLysProAlaThrLys 785
||||| : : : : : ||| : : : : : ||| : : : : :
2901 AAA...GCTCCTGCTGCAACTCAA 2921

seq_name: gb_pat:A93836

seq_documentation_block: 2766 bp DNA PAT 22-JAN-2000
LOCUS A93836
DEFINITION Sequence 5 from Patent W09728263.
ACCESSION A93836
VERSION A93836.1 GI:6741967
KEYWORDS
SOURCE Bacillus stearothermophilus.
ORGANISM Bacillus stearothermophilus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Geobacillus.
REFERENCE 1 (bases 1 to 2766)
AUTHORS Lubitz W. and Sleytr U.
TITLE RECOMBINANT EXPRESSION OF S-LAYER PROTEINS
JOURNAL Patent: WO 9728263-A 5 07-AUG-1997;
LUBITZ WERNER (AT); SLEYTR UWE (AT)
FEATURES
source
1. 2766
/organism="Bacillus stearothermophilus"
/strain="pv72"
/db_xref="taxon:1422"
1. 93
sig_peptide
1. 2766
/feature="unnamed protein product"
CDS
1. 2766
/codon_start=1
/transl_table=11
/db_xref="GI:6741968"
/translation="MAYQPSFRKFEVATTAIVASAVAPVSAASFVDVAPQYKDA
IDPLVSTGATFKKTEKFCGYDEITRLDAVILIARLYKLDONAKDAGPTDVRDRAK
YVNAIVEAVLVNGKARGKRCAGADPLIRVEMAKIIRYKLDQVRLPPTDNDMPAP
YKALIKTEVTRKLRHQASVHKNTITLDFPOFVIRAVNLNAVPEIYEVYAVNSTTV
KVFENQIADVDFTNFADNGLTVTKATLSRQKSEVVVNNPFTNDEYTTATGIRK
NKGETAKELETGEFVWSYODAVTVALLNSLSRKGESGLTVDDGKDVYGAKEELTS
SNNTIYVSSGEVSAAKVTAVKPGTADVTAKVTLPDGVVLTNFEKVTYVAPVOV
NOGFTLVDLNSAPONTVAFNAEKVTSMEAGETKVAVYDRKNGDEPKPVDKDAT
VSLNPITITTAALINGSELLVTANAGSGKASEPVLTKDKTKFTFYDVKKDDVIDIK
VDAITSYKLSDEAVGGEGVEGCVNKKITIKVASVDQYGEKIRFTRKATVTTNBLGVIK
MNSDNTIDFDSGNSATDQFVVAVTRDKLVNKEVEKYRKASDTPISFKITIVAVV
NYKADATPVGLDIVAPSKIDVNAIPNTASTADVDIFNESEVETITLDSNGRQRKYPT
ATTIVGTRKKKRVNGVLOFGKNEELTSTSSGVNCGTACGMRIRIGKXINSAYV
PASATVATSPVYKLVNSDNDLFEELIGVIDPQVLKDEIDINERIVASKAKNDGY
LYNKPPLYTVYKDSAGKYIRPGANCANYGICNHATNCGNIMPDEQGLAKKPSDVHEDNES
LTNVKATGSGTVSSPSLSDAIQLTNSGAVSFTLVIKSIYKAGADKDNLLIAPVS
VNVTVYRK
94.. 2763
mat_peptide
94. 2763
BASE COUNT 996 a 465 c 534 g 771 t
ORIGIN
alignment_scores:
Quality: 419.50 Length: 879
Ratio: 0.891 Gaps: 40
Percent Similarity: 53.584 Percent Identity: 24.005
```


BASE COUNT	972 a	463 c	538 g	790 t
ORIGIN	mat_peptide 94..2760 /product="unnamed" /codon_start=1 /transl_table=11 /protein_id="CAB77070.1" /db_xref="gi:7240687" /translation="MAYQPSRKRFVATTATAAMVASAVPVVSSAASFTIDVAPQYKDA IDFLVSGATGKTEKEGVYDELITRLDAAYILARVILKLDNAKDAGDFEVPDRAK VYMLAEVGLNGKAPGKGFAYDPLTRDEMAKILANRYKJADADVKTLPFVNTDTPAK VYRLKLYEYTKGKTPPSFGAYQNTITGDEPAKIVANVAPEIYEVAVNSTYTK VTEPQTQIADVDFTFPAIDNGLTJYTKILSRKSEVYVAKPEPTRNDEYITTAGINK LKGTAELELTGEVWSAQVADAVTALNNLSLTVGEESGJTKADQGVKVGAKLETSS NFNIVVSSGCEVSAKVYVAVPGTADYAKVILPGVILTNFKYVLEVPAQVON QGFVLVNLNSAIPONTVYANAEKVTSFNAEETVYAKYDKNDEPEKPEVDFADAVV RSLNPILATIAIPONTSELTVTANAGCSGSAEYVFKNTKFTFVYDKPEVLDQIV DATSKLTDEAGVEGVENQKTIKSAVOQYKEKLEFGKAVYATNTNEGIVIKRKK VNSNTITDDESGVATDOOFVYATKRVKVCVKKYFKKNSDTPSTKTIIVNVN VKALATPYGLDIYAPSELDVNAVNPASTADYDFINFSVELYITLDSGNRLKTKPYAR TTLVGTNDIYEVNSVLOFGKNDLELTLLTSSITVNDVATPAGITRILPYKTIKNSAVP ASAVIYASPVYTKLNNSSNDLTJEEELTEGVLDTPQVAKDEIDNEFLASKARNDGIL VNKPLVYIKDASGRVILPGANVYGLINDATNGNTWPEEQGLAKKFSDFVHFDVFSIL ANVYKGTSGTVSSPSLSDAIQLTNSGDAVSFTLIIKSIYKAGADKDDNNLLAPVSN NVYVTK"			

```

alignment_scores:
    Quality: 415.00      Length: 846
    Ratio: 0.918        Gaps: 34
    Percent Similarity: 53.428    Percent Identity: 23.995

alignment_block:
US-09-754-947-1 x AX000222 ..

Align seg 1/1 to: AX000222 from: 1 to: 2763

4 ThrPheProAspValProAlaAspHisTPrGlyIleAspSerIleAsnTry 20
:::||||| ||||| ::::: |||::|::|::|::|
97 AGCTTCACAGATGTTCGCCGCGCAATAT....AAAGATGCGATCATTTT 140
rIeuValIGluLysGlyAlaValAlaLysGlyAsnAspLysGlyMetPheGlu 37
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
141 CTTAGATATCACTGCTGTCACACAAAGGTAAACAGAAACAATATTCGCCG 190
37 rGcLYLysGluLeuThrArgAlaGluAlaAlaThrMetCetaIaGlnIle 53
|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
191 TTTCACATGAAATCACTGCTGTAGATGCGCGCATTTTCTTGCACAAAGTA 240

```


383 TAAACGTTAAAGCACTGGCAAAATTGGTGCATACGACCCATTAACTCGC 432
103 ValSerMetAlaSerLeuLeuValGluAlaTyrLysLeuAsp..... 116
433 GTTGAATAATGGCAAAATTCATCGGGAACCGTTACAAATTTAAACCTGACGA 482
117ThrLysValAsnGlyThrProAlaThrLysPheI 128
483 TGTAAACATTCATTCATGATGATTAACGATACATGGCACCATTACGTAA 532
128 ysaSplLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
533 AACCGCTTTATTAATACGAAGTAACAAAGGTAA..... 567
145 GluLeuGlyIleSerValGlyThrGlyAspGluTrpGluProLysLysTh 161
568ACACCACACACCTCGGTGCATACCAAAA 596
161 rValThrLysAlaGluAlaGlnPheIleAlaLysThrAspLysGluP 178
597 CATCACATCGCGGTGACTTTGGCAATTTGGATATAGACGGGTGAATATT. 645
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrGln 194
646AATGCAGTGCAGAAATAGTTGAAGTAACCTCGGTAAATTCGACT 690
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211
691 ACAGTGAACCTAACATTCATACGCCAATTT.....GCTCA 725
211 PileLysValThrAsnLysAlaAsnAspLysValLeuValLysGluV 228
726 TCTTGTATTTCACAAAATTTGCTATCGATTAACGTTAACTGTTACAAAG 775
228 alThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsn 244
776 CAACCTTTCTCGTGTAAATAAATCCGTAGAGTTGTGTAAATAAACCG 825
245 LeuAlaAlaIleGlnThrTyrThrValAspValAsnLysValGly..... 259
826 TTACTCTGTAATCAGCAATATACAAATTACACGACAGCATTAATAAATT 875
260 LysThrGluValAla.....ValGlySerLeu. 268
876 AAAAGCGCAGACCGGTACGAATTAACGTGTAAGTTGTTGGTGTTC 925
269 ..GluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspGlu 284
926 AAGATCGCGTAACCTGTGCATTAATTAATGATTCCTTAAAGTTGGAGAG 975
285 ProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValVa 301
976 GAATCTGGTTTA.....ACTGTAAAGATCAGAGTGGCAAAAGATGTGT 1019
301 L.....SerProGluGlyIleGluPheValT 310
1020 ACGTGTAAAGTAGAAGTACTTCTCTTAATACTAATATTTGTTAGATT 1069
310 hrProAlaAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
1070 CAAGTGGCCAAATATCAGTATCTGCTGTAAAGTTACAGCTGTAAACCG 1119
327 GlyThrSerThrThrValLysAlaValTyrLysAspGlyLysValVa 343
1120 GGAACACGCTGATGTTACTGCAAAAGTTACATTAACAGATGGGTGTACT 1169
343 AlaGluSer.....LysGluValLysValSerAlaGluG 355
1170 AACCAATATCATTTAAAGTAGACAGTACAGAACTGCTGTGCAAGTACAA 1219
355 LysAlaValAlaSerIleSerAsnTrpThrValAlaGluGln..... 369

615 LeuLysValThrThrAsnLysGluGlyLysGluValAspAlaThrAs 631
: : : : :
2092 ACTACACTGTAGACTAAT.....GATATGTTGAAGTAATATGC 2132
631 PAlaGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGlyA 648
: : : : :
2133 GAATGATTAATCAATTCAAGGTAACGATGAATTAACGCTA.....T 2173
648 IalysAlaGluGluThrThrLysValThrValIalleuAspGlyLysLeu 664
: : : : :
2174 TAACTCTCTAGTACGTAACGTTGATGTAACGCTGATGGA..... 2217
665 IleThrThrHis...SerPheLysValIalAspThrAlaProThrAlaLys 680
: : : : :
2218 ATTACAAACGCTATTCAGTAAATAATATCAACTCGCAAGCTGACGTCG 2267
680 sGlyLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProA 697
: : : : :
2268 CAGTGCACACAGTACGACACAGTCTGTTACTGTTAAGCTTAATTCAGCTG 2317
697 snAlaAspLeuLysAlaIalaleuAsnIleLeuSerValAspGlyVal 713
: : : : :
2318 ATATGATTTAATCACTTGAAGATTAATATTCGCTGTAATTCAC..... 2361
714 ProAlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerAl 730
: : : : :
2362 CCTACACAAATTAGTCAAAGATGACACATCAAC...GAATTTATTCAGCT 2408
730 asPheThrAsnValIalAlaGluAsnGlyThrValGlyValLysGlyAlaT 747
: : : : :
2409 T.....TCAAACGGCTAAATAAATGATG 2431
747 hrSerIleThrValLysAsnLeuThrValIalLysAsp 759
: : : : :
2432 GATATTTGTAATTAACCGCTGTTAAGCTTAAGAT 2469
seq_name: gb_pat:AX033305
seq_documentation_block:
LOCUS AX033305 2766 bp DNA PAT 21-SEP-2000
DEFINITION Sequence 3 from Patent DE19903345.
ACCESSION AX033305
VERSION AX033305.1 GI:10280105
KEYWORDS
SOURCE Bacillus stearothermophilus.
ORGANISM Bacillus stearothermophilus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Geobacillus.
REFERENCE 1 Lubitz,W.
AUTHORS Patent: DE 19903345-A 3 03-AUG-2000;
JOURNAL LUBITZ WERNER (AT)
FEATURES
source location/Qualifiers
1..2766 /organism="Bacillus stearothermophilus"
/db_xref="taxon:1422"
1..93 /protein_id="CAC09983.1"
1..2766 /db_xref="GI:10280106"
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09983.1"
/db_xref="GI:10280106"
/translation="MAYOPKSYRKFVATTTAAMVASAVPVSAASFQDAPQYKDA
IDELVSTGATKCKTEKFGYDEITRLDAVILARILKLDVNAKDAAGFTDVKDRK
YVALVAGVLAASKAPKRGAYDPLRVEMAKITANRILKADADKLPFDVNDITAP
YVALKYEVTKRLKHOQASVHRKNITLRDFAQVYRAVINAVPEIVETAANSTTV
KYTFNIOIADVDFTNFAIDNGLVTRKATLSRDKSYEVVNVKPFTRQEVTTATGK
NLGKTAKELTGKFMVSODAVVALNLSLKGESGLVVDODGKDVGAAGVELTS
SNTNIVVSGEVSAAKVTAVKPGADVTATVTLDPGVLLNTEFKVTAEVPOVO
NGFTLVNDLSNAPORTVYAFNKAEKVTSMACETKRYAMDTKNGDEPKTPVDFKAT
VRSNLPITATINSGSELVTANAGOSGAKSFEVTRKNTKRTFYDVKADPVLQDTK

mat_peptide /product="unnamed"
BASE COUNT 973 a 464 c 537 g 792 t
ORIGIN
alignment_scores:
Quality: 413.50 Length: 846
Ratio: 0.913 Gaps: 35
Percent Similarity: 53.546 Percent Identity: 24.350
alignment_block:
US-09-754-947-1 x AX033305 ..
Align seg 1/1 to: AX033305 from: 1 to: 2766
4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
: : : : :
97 AGCTTCACAGATGTTGCCCGCAATAT.....AAAGTGCAGTGAATTT 140
20 rLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluP 37
: : : : :
141 CTTACTATCAACTGCTGCACAAAGGTAAACAGACAAATTCGCGC 190
37 rGlyLysGluLeuThrArgAlaGluAlaIalThrMetAlaGlnIle 53
: : : : :
191 TTACAGATGAATCATCTGCTAGATCGGACGATATCTTGCAAGATA 240
54 LeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGl 70
: : : : :
241 TTAACATTAAGAGCTTGACACGCAAAAGACGACGCTTCACAGATGTC 290
70 nGlyInTrpTyThrProPheIleAlaIalValGluLysAlaGlyAlaI 87
: : : : :
291 AAAGAC...CGTCAAAATACGTCAACGCCGTTGTAAACGTGCGAT 337
87 IeLysGlyThrGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102
: : : : :
338 TAAACGTAAAGACACCTGCAAAATTTGGTCATACGACCAATTACTGCG 387
103 ValSerMetAlaSerLeuLeuValGluAlaIalThrLysLeuAsp..... 116
388 GTTGAATGGCAAAATCATCGCGAACCCTTACAAATTAAGCTGACGA 437
117ThrLysValAsnGlyThrProAlaThrLysPheL 128
438 TGTAAACTTCATTCACCTGATGATTAACGATACAGGACACCATACGTA 487
128 ysAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
488 AAGCGCTT.....TATAATACGAAGTAAACCAAAAGGTTAAAA 525
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProLysLysTh 161
526 CACCAACAAGCTTCGTCGTCATAC.....AAAA 554
161 rValThrLysAlaGluAlaIalGlnPheIleAlaLysThrAspLysGluP 178
555 CATCACTTCGCTGACTTGGCGCAATTTGTATATAGCGGATGATAT. 603
178 heGlyThrGluAlaLysValGluSerAlaLysAlaValThrGln 194
604AATGACAGTGCAGAAATAGTTGAAGTACGCGGTTAATTCGACT 648
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211

649 ACAGTGAAGTAACTCAATACGCAAAATT.....GCTGA 683
211 PileLysValThrAsnLysAlaAsnAsnLysValLeuValLysGluV 228
684 TGTGATTTCACAAATTTTCTATCGATACGTTAACTGTTACTTAAG 733
228 aLThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsn 244
734 CAACCTCTTCGCTGATATAAAATCCGTAGAGTGTGTTAAATAAACCG 783
245 LeuAlaLeuLysGluThrTyrThrValAspValAsnLysValGly.... 259
784 TTACTCGTAATCAGAAATATACAAATTACAGCGACGCAATTAATAATTT 833
260 LysThrGluValAla.....ValGlySerLeu. 268
834 AAAGCGGACCGCTAAGCAATTAAGTAACTGTTGTTGCTGCTTC 883
269 ..GluAlaLysThrIleGluMetAlaAspGluThrValAlaAspGlu 284
884 AAGATCCGCTAAGTGTGCACTAAATATAGTTCGCTTAAGTTGGAGAG 933
285 ProThrAlaLeuGluPheThrValLysAspGluAsnGlyThrGluValVa 301
934 GAATCTGGTTTA.....ACTGTAAGATCAGAGTGCAGAAAGATGTTGT 977
301 L.....SerProGluGlyIleGluPheValT 310
978 AGGTGCTAAGTAGAAGTAACTTCTTCTTAATCTAATATTTGTTGTAAGTT 1027
310 hProAlaAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
1028 CAAGTGGCCAGACTATCCTGCTGCTAAAGTTACACCTGTAAACCG 1077
327 GlyThrSerThrThrValLysAlaValTyrLysLysAspGlyLysValVa 343
1078 GGAACACGCTGATGTTACTGCAAAAGTTACATTACAGATGGTGTGTACT 1127
343 AlaGluSer.....LysGluValLysValSerAlaGluG 355
1128 AACAAATACATTTAAAGTGACAGTTACAGAACTGCCTGCAGAAAGTAA 1177
355 LysAlaAlaValAlaSerIleSerAsnTyrPheValAlaGluGln..... 369
1178 ATCAAGGATTTACTTAACTTGAATCTTCTTAAGTCTCCACAGAAATACA 1227
370AsnLysAlaAspPheThrSerLysAspPheLysGluAsnAs 383
1228 GTTGCATTTACAAAGCTGAGAAAGTAACTTCAATGTTTGTGAGAAAC 1277
383 nLys.....ValTyrGlu.....GlyAspAsnAlaTyrValG 394
1278 TAAACACATGTCATGTATGATACATAAAACGCTGATCCTGAAGTAAAC 1327
394 LysValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGlu 410
1328 CTGTGATTTCCAAAGAT.....GCAACT 1350
411 TyrGluSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
1351 GTACGTTATTAATCAATTAATTCACACAGCTGCTAATTAATGTAAGTGA 1400
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
1401 GCTCCTCTCACAGCTAATGCTGCCAATCTGGAAGACCTTCATTGGAAG 1450
440 aLThrValLysAspSerLysGlyLysAlaLeuValSerHisThrValGlu 456
1451 TAACATTTTAAGATTAAT.....ACAAAGAACATTTTACA 1485
457 IleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473

1486 GTTGATGTGAAAAAGACCCTGTATTACAAAGATATTAAAGTAGATCAAC 1535
473 rAsnValAlaLeuSer.....ThrLysAspV 482
1536 TTCTGTTAACTTTCCAGTGAAGCTGTGGCGGGGAAGTTGAAGAG 1585
482 aLThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGlu 498
1586 TTAACCAAAAAAGATTAAGTAAGTGCAGTTCACCAATACGTTAAGAA 1635
499 phe.....ThrAlaPr 502
1636 ATTAATTTGTGACAAAAGTAAAGTACTTGTACAACTAATATACAGAACG 1685
502 oValThrValLysValLeuAspLysAsp.....G 512
1686 ACTAGTATTAAATAATGTAAATAGCATATATCAATTGACTTTGATAGCG 1735
512 LysGluLeuLysGluGln..... 518
1736 GCAATAGTCACACTGACCAATTTGTTGCTTGCAACAAAGACAAATTT 1785
519LysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAs 532
1786 GTCAATGTAAAGTAGAAGTAAATATTTCAAAAT..... 1821
532 nAlaAlaGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLys 549
1822 ...GCTAGTCACACACCACTCACTCAAAATTAATCTGTTAAG 1867
549 eGlyGluLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1868 TAGTAATGTAAAGCTGACCTGACACAGTAGATTAATATGATTTAGTACA 1917
565 ProGlyAlaPheSerLysPheGluValArgLysLeuAspThrGluLeuAs 581
1918 CCT.....TCTGAATTTGATGTGAATGCTCCAAACACT..... 1950
581 PLYSTYrValThrGluGluAsnGlnLysAsnAlaMetThrValSerValL 598
1951 .GCTTCTACTGACAGATGTTGATTATTATTTTCGAAGGTGTGAGATT 1999
598 euProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu 614
2000 ATACACGATCTAATGATGTAACCTCTTAAAGAACTTACCAACTGCA 2049
615 LeuLysValThrThrThrAsnLysGluLysGluValAspAlaThrAs 631
2050 ACTACACTGTGAGTACTAAT.....GATTATGTTGAAGTTAAATGG 2090
631 pAlaGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGlyA 648
2091 GAATGATTTTACAATTCAGAGGTAACGATGAATTAACGCTA.....T 2131
648 lAlysAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLeu 664
2132 TAACCTCTTCAAGTACAGTAAACGTTGATGTACAGCTGATGGA.... 2175
665 lIleThrThrHis...SerPheLysValAlaAspThrAlaProThrAlaL 680
2176 ATTACAAAACGATTTCCAGTAAATATATATCACTGTGCAAGTGAAGTCC 2225
680 sGlyLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProA 697
2226 CAGTGCACAGTACAGACAGTCCCTGTTACGTTAAGCTTAATTAATGATG 2275
697 snAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyVal 713
2276 ATAATGATTTAAACATTTGAAGAAATTAATATTCGGTAAATGAC..... 2319
714 ProAlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerAl 730
2320 CCTACCAATTAATGTCAAAGATGAAGACATCAAC...GAATTTATTTGCACT 2366

```
730 aaspthrsnvalvalaagluasnclYthrValglalalysglYalaT 747
      :          :          :          :          :
2367 T.....TCAAAGCGCGCTAAAAATGANG 2389

747 hserlleyVallysasnleuthrValVallysasp 759
      :          :          :          :          :
2390 GATATTGTATATAAACCGCTTAAACGTTAAAGAT 2427

seq_name: gb_ba:BACSLPZ

seq_documentation_block:
LOCUS      BACSLPZ               4251 bp      DNA      BCT      03-OCT-1994
DEFINITION Bacillus sphaericus 2362 (clone pgC4.2) surface-layer protein gene,
complete cds.
ACCESSION  M28361
VERSION    M28361.1 GI:341911
KEYWORDS   surface layer protein.
SOURCE     Bacillus sphaericus (strain 2362) DNA.
ORGANISM   Bacillus sphaericus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/staphylococcus group; Bacillus.
            1 (bases 1 to 4251)
REFERENCE  Bowditch,R.D., Baumann,P. and Yousten,A.A.
            Cloning and sequencing of the gene encoding a 125-kilodalton
            surface-layer protein from Bacillus sphaericus 2362 and of a
            related cryptic gene
JOURNAL   J. Bacteriol. 171 (8), 4178-4188 (1989)
MEDLINE   89327128
FEATURES   source
            1..4251
            /organism="Bacillus sphaericus"
            /strain="2362"
            /db_xref="taxon:1421"
            -35_signal
            -10_signal
            RBS
            CDS
            164..168
            178..3708
            /codon_start=1
            /transl_table=1
            /product="surface layer protein"
            /protein_id="AA50256.1"
            /db_xref="GI:556012"
            /translation="MAKONKGRKFEFAASATAALVASAIVPVASAAQINDENKISGVAK
            EAYOSLIDAGVIOGDANGFNPLKTIISRAEATITFTALAEAGDVFNDKADAVY
            YDAIAATVENGIEFGVATGEFAPKQJLRSEAKIIVDAFELBEGDLSERADASTVK
            PMAKSYLEIVANGVIGSEANGKTNLNPAPITROPFAVFSRTINVAATPKVDKI
            EYDAKALNTLSDGKETVTELEKALPNKETEYTRIKOYEYAKTYVYVTTATK
            SVSATNLKEVYVEPDGTVDKETADANNTALSKSRTIKSVSLADNKTATVTTDLKLN
            NNRKADALISINVAKGDEIIVKNVEFAVDNKIPEVIEVSLGTRAKAVTLSEVENL
            SSTNFTLDGKAYFGNVVAGANKTVILTPYSSSLVSGDKLTVSGAKDFAGKEVLSN
            THEKVVEDKEAPVTEATATLETVTLEFSEDIDMDIYKSNVNMKGSGDSKEASEE
            RIDNKKYFVEKSEKTLPTGKVDVYVEDIKDYSDNKLADTKTATVPEIDQTPREYR
            KVTALDEKTIKVTESKTVDESAIKTGNVYKXKDDKVVSDKTYVDSKSKSYITDL
            YSKVSGEENTTTIKNVADATKLNNTMLDTGKTRSDKEGDIYEHVNAEDKAKKVV
            KFDKMDAASLADSNLVKINDTLOLSEDAVLVSNATVVTITFAETIKADVVV
            FASGKAISGSKVNVNELQWGVKDTSGNVHKGSGENKITLSTSTPLAKITDKD
            YDAKYTAELVDKRTYKVFSTVINSAAANFTSESHKIDSIOVNGSTVYKFEDEIN
            TMSDDLKVLNLSKLVDIAGNESTNTPAIKAGINDLSVAAPVGEPPVDKETIFE
            TFSNLTSSVSGEVLSPDPTVRYSDNKLAIKRYSAVAINNNNVVITLSDNREVAA
            YKTRAKNAKLITDNGDKNAIADFTTKTATKVASGTLSDAKKTMINNETRAKKA
            KATGTEGTAANTNOIVGSKDALQVAIDVAELVKNPTAATLOOLDAKTDLVATITAYNA
            AKVEDISLVAADPLVGLTNDNGTITGFVAGTGTLLKVTSDSAANVEVTDPTGLAVTA
            KAKGEANILVOVLTGKDKVYIKGVKVTYVSE"
            3764..3806
            /note="putative"

terminator
BASE COUNT  1552 a      681 c      781 g      1237 t
ORIGIN
```

```
alignment_scores:
Quality: 402.50      Length: 916
Ratio: 0.833         Gaps: 36
```

```
Percent Similarity: 52.729      Percent Identity: 23.144
alignment_block:
US-09-754-947-1 x BACSLPZ ..
Align seg 1/1 to: BACSLPZ from: 1 to: 4251

16 AspSerlleasntrYleuValglulysglYalaValLysglYasnAsply 32
      :          :          :          :          :
310 GAAGCAGTTCAATCTTAGTACCGCTGTGTAATCCAAAGGAGGCTTAA 359

32 sglYmetPhegluProglYlysglulLeuthrArglaglalaalatrTM 49
      :          :          :          :          :
360 CGGCACCTTCAACCCACTTAACGTAATCTACGTCAGCGGAGCTGTACAA 409

49 etMeLaglInlleuasnleuProlleasplyAspAlaLysProser 65
      :          :          :          :          :
410 TCTTCACATACCGCTCTAGATTAAGACGAGAGTGATGTA.....AAC 453

66 PheAlaAspSerGlnGlyGln...TrpTyrThrProPheIleAlaLava 81
      :          :          :          :          :
454 TTCAAAGACGTTAAAGCGATGCTGTGACTACGATGCTATCGCAGCAAC 503

81 lglulysAlaGlyValIlelysgly...ThrglysnGlyPhegluProA 97
      :          :          :          :          :
504 TGTAGAAACGGAAATTTTGAAGGTGTAACTGTACTGATTCGACACCAA 553

97 snglYlysIleasprqValSerMetAlaSerleuLeuValglulAtry 113
      :          :          :          :          :
554 ACAACAATTAATCTGCTGTCTGAGCTGCTAAATTTTACTAGATGCTTTC 603

114 LysLeuAspThrLysValasnGlyThrProAlaThrLysPheLysAsple 130
      :          :          :          :          :
604 GAATTAGAGCGTGGAAGCGCAT.....CTACAGCAATTCGTGCGACGC 644

130 uclurhrlLeuasn...TrpGlylysglulysAlasnIleleuValglul 146
      :          :          :          :          :
645 TTCTACTGTAAACCATGGGCTTAATCTTACCTAGAAATCGCAGTTGCAG 694

146 euGlYIleSerValGlyThrGly..... 153
      :          :          :          :          :
695 ACGGCGTTATCAAGGTTCTGAAAGCAATGTTAAACCAATTAAACCCA 744

153 ..... 153
745 AATGCTCAATTACTCGCCAGACTTCGAGTTGATTCACGACTAT 794

154 ..... 157
795 TGAAGACGTAGATGCTACTCCAAAGTTGACAAATCGAAGTAGTTGACG 844

158 ..... 163
845 CTAAGAACTTAAACGTTACTTATCTGACGCTACGTAAGAAACTGTTACT 894

164 LysAlaGluAlaGlnPheIleAlaLysThrAspLysGlnPheGly.. 179
      :          :          :          :          :
895 TTAGAAGAAAGCTTAGAGCCTTAACAAAGAAACAGAAAGTTACTTCAAAAT 944

180 ..... 183
945 TAAGCATGTTGAATACAAAGCTAAAGTTACTTATGTTGTAAGTACGTA 994

183 lalysValgluSerAlaLysAlaValThrThrglnLysValglulValys 199
      :          :          :          :          :
995 CTCACGATTAATCTGATCTGACACATACCTTAAGAAATAGTAGTTGAA 1044

200 PheSerLysAlaValglulysLeuthrLysglulAspIleLysValThrs 216
      :          :          :          :          :
1045 TTCAGACGTACTGTGATTAAGAAACAGCTGAAGAT.....GCAGCTAA 1088

216 nlysAlaAsnAspLysValleuValLysglulValThrLeuSerGlnA 233
      :          :          :          :          :
I      III      :          :          :          :          :
I      III      :          :          :          :          :
```

1089 CTACGCTTAAAAATCAGGTAAACAATTAAATCTGATCTTACGCTG 1138
 233 sPLysaIrgSerAlaThrValGluLeuTySeraIleuAlaIalysGln 249
 ||| : : : : : ||| : : : : :
 1139 ATATATAACAGCTACTGTTACACTTAACGATTAACCTTAACAACAA 1188
 250 ThrTyThrValAspValAsn.....LysValGlyLysThrGluValAl 264
 :
 1189 GCGAATGCTATTAGCATTTCTTAATGTAAGCGCGCATMAAGAAATCAA 1238
 264 aValGlySerLeu..... 268
 ||| : : : : :
 1239 TGTGAAAAATGTGAATTACACAGCTAGATATAAATCCGAAGTTA 1288
 269 ..GluAlaLeuThrIleGluMetAlaAspGlnThrValAlaIalysGlu 284
 ||| : : : : : ||| : : : : :
 1289 CTGAAGTTAAATCTTAGMAACAAAGCTGTAAAGTTACATTATCTGAA 1338
 285 Pro.....ThrAlaLeuGlnPheThrValLlysAspGlu..... 295
 ||| : : : : : ||| : : : : :
 1339 CCGTTGACAACTTAAGTCTTACAAACTTCACCTTAGACGGAAGCGTA 1388
 296AsnGlyThrGluValValSerP 303
 ||| ||| ||| : : : : :
 1389 CTTGGGAATGTTGTTATGGCTCTGCTAATAAACAGTTATCTTACAC 1438
 303 roGluGlyIleGluPheValThrProAlaIalGluLysIleAsnAlaLys 319
 || : : : : : ||| : : : : :
 1439 CTTACAGCTCTTCTCGCACTTCTGTTGCGGACATCAAGCTTACTGTTTC 1488
 320 GlyGluIleThrLeuAlaLysGlyThrSerThrThrValLysAlaValTy 336
 ||| : : : : : ||| : : : : :
 1489 GGTGCAAAAGACTTCGCTGTTTGTTCATTAATTAATCAACGCATGAAT 1538
 336 rLysLysAspGlyLysValValIalGluSerLysGluValLysValSera 353
 ||| : : : : : ||| : : : : :
 1539 TAA.....GTTCTGACAGATMAAGAACCAACCAAGCTAA 1573
 353 IalGluGlyAlaValAlaSerIleSerAsnTrpThrValAlaGluGln 369
 :
 1574 CTGAAGCTACTGCACCTTGAAACCTGTACATTACATTCTCAGAAAT 1623
 370 AsnLysAlaAsp.....PheThrSerLysAs 378
 ||| : : : : :
 1624 ATTGACATGATCTCTAAAGCTTCAACGTTTATTGGAATCTCGAAG 1673
 378 PheLysGlnAsnAsnLysValTyTrGlu.....GlyAspAsnAlaTyTr 393
 ||| : : : : : ||| : : : : :
 1674 TTCTAAGAAAGAGCATCTGAATTCGAGCGTATTCGCTGATTAATTA 1723
 393 aIGlnValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysVal 409
 ||| : : : : : ||| : : : : :
 1724 AATTCGTA...TTCAAAAGCTCTGAAACAACTCTCCAACTGGAAGTA 1770
 410 GluTyTr.....GluSerLeuAsnThrGluValAlaVala 420
 :
 1771 GATGCTACGTGAAGACATTAAAGTTACTCTGATACAAATGCTAA 1820
 420 lValAspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaP 437
 ||| : : : : : ||| : : : : :
 1821 ACATCAAAAGTTACTGTAACCTCGTAATCGCAAACTCGTCCGAAG 1870
 437 roValLysValThrValLysAspSerLysGly...LysAlaLeuValSer 452
 ||| : : : : : ||| : : : : :
 1871 TAAAGAAAGTAAGTCTCTTGAATGAAAAAATCATCAATCTCTCT 1920
 453 HlStrValGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleTy 469
 ||| : : : : : ||| : : : : :
 1921 AAAACTGTTGAGAGAGCTCTGATAAAACAGCAACTACACAGTTAA 1970
 469 sLeuGluLysThrAsnValAlaLeuSerThrLysAspValThrAspLeu 486
 ||| : : : : : ||| : : : : :
 1971 AGACAAAGCAGCAAGTAGTTCTGTTGATTAAGTAACCTGTTGATTCAA 2020

486 ys...ValLysAlaProValLeuAspGlnTyTrGlyLysGluPheThrAla 501
 ||| : : : : : ||| : : : : :
 2021 AAGATTCTAAATCTGTAATATTATGACCTATACCAAAAGTAAGTGTGG 2070
 502 ProValThrValLysValLeuAsp...LysAspGlyLysGluLeuLysG 517
 ||| : : : : : ||| : : : : :
 2071 GAAAAATCAATTTACAAATTTAAGATGTTAAAGATGCAACAAACCTTA 2120
 517 uGlnLysLeuGluAlaLysTyTrValAsnTrpGluLeuValLeuAsnAla 534
 :
 2121 TACAAATGCTAGAT.....TACAGTGTAA.....TTTACAGAT 2155
 534 IalGlyGlnAlaGlyAsnTyTrThrValValLeuThrAlaLysSerGly 550
 ||| : : : : : ||| : : : : :
 2156 CAGATTAAGAGAGCTCCAGATTTATGACAGTATCAATGCTATGCA... 2202
 551 GluLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAl 567
 ||| : : : : : ||| : : : : :
 2203 ...AAAGCTAAAAAGTTGTTTAAATTCGACAAAAAATGATGATCAGC 2249
 567 a.....PheSerLysPheGluValArgGlyLeuAspThr..... 578
 ||| : : : : : ||| : : : : :
 2250 ATCTTAGCTGACTATTCAACTACTAGTAAAAATCAATGATACTTTAC 2299
 579 ..GluLeuAspLysTyTrValThrGluGluAsnGlnLysAsnAlaMetThr 594
 ||| : : : : : ||| : : : : :
 2300 AAACATTATCAGAAAGATGTTGCTACACTTCAAGTTCAATATGATGCTACA 2349
 595 ValSerValLeuPro.....ValAs 601
 ||| : : : : : ||| : : : : :
 2350 GTAGTACATTAATCTTTGACAGACAAATTAAAGTGACAGATGTTTAAT 2399
 601 PalAsnGlyLeuValLeuLysGlyVala.....GluAlaAlaGlu 615
 ||| : : : : : ||| : : : : :
 2400 TGCTTCTGTAAGCAATTTCTGATCTGCTAAGTAAGTAATGTAAGAT 2449
 615 euLysVal.....ThrThrAsnLysGluGlyLysGluVal 627
 ||| : : : : : ||| : : : : :
 2450 TACAAATTTAGGAGTAAGATATCTTCTGTAACGTGCAATGAAGAAATTT 2499
 628 AspAlaThrAspAlaGlnValThrValGlnAsnAsnSerValIleThrVal 644
 :
 2500 AATGTTCAGAAATTAATAATTTACTTTATCATCAACATCA.....ACGCC 2543
 644 lGlyGlnGlyAlaLysAlaGlyGluThrTyTr.....LysValThrVal. 658
 ||| : : : : : ||| : : : : :
 2544 ATTAAGACTGCTAAATGACAAACGCTTCTGATTTGATTTAAAGTAAT 2593
 659 ..ValLeuAspGlyLysLeuIleThrThrHisSerPheLysValValAsp 674
 :
 2594 AATTAGTTATGAAAGAACTGTAAGATTAATTTCTACACTTATTTAAC 2643
 675 ThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLe 691
 ||| : : : : : ||| : : : : :
 2644 TCAGCA.....CTGCAAAATGCATTTACTTCTGAAATCA 2678
 691 uLys..... 692
 ||| : : : : :
 2679 TAAATTTGATTTCTATTCAGTGATGATGCTCAACTGTTACAGTTAAT 2728
 693GluValAlaProAsnAla.....AspLeuLysAlaAla 703
 ||| : : : : : ||| : : : : :
 2729 TCAAGATGAAATTAACAAACGCTTCTGATTTGATTTAAAGTAAT 2778
 704 LeuLeuAsnIleLeuSerValAspGlyValProAlaThr.....ThrAl 718
 ||| : : : : : ||| : : : : :
 2779 TTATCTAATTAATTTGATTTGACAGAAAGCAAGTACAAATTAATATCC 2828
 718 aLysAlaThrAlaSerAsnValGluPheValSerAlaAspThrAsnValTy 735
 ||| : : : : : ||| : : : : :
 2829 AATGCCAATTAAGCTGATTAATTAATTTAGATTCTGTTGACCAAGTTG 2878

```
735 alalaglansnclgthrvalglvalalysglyalathrserletyrrval 751
|||||  ::::  |||  ::::
2879 ttgttagagagccctgctgttataaagaacatattccttcttcca 2928
752 lysasnleuthrvalvallysaspglysgluglnlysvaigluphe 767
:::|||||  |||  ::::  |||  ::::
2929 gaaacttaactagctgttcttattgagaaagttttaagacagacttt 2976

seq_name: gb_ba:AF068060

seq documentation block:
LOCUS AF068060 4889 bp DNA BCT 31-Dec-1999
DEFINITION Cytophaga sp. 'Jeang 1995' S-layer protein precursor, gene,
complete cds.
ACCESSION AF068060
VERSION AF068060.1 GI:6648626
KEYWORDS
SOURCE .
  Cytophaga sp. 'Jeang 1995'.
  Cytophaga sp. 'Jeang 1995'.
  Bacteria; CFB group: Cytophagales; Cytophagaceae; Cytophaga.
REFERENCE
  Jeang,C.L., Lee,Y.H. and Chang,L.W.
  1 (bases 1 to 4889)
  Purification and characterization of a raw-starch digesting amylase
  from a soil bacterium--Cytophaga sp.
  Biochem. Mol. Biol. Int. 35 (3), 549-557 (1995)
JOURNAL
  Biochem. Mol. Biol. Int. 35 (3), 549-557 (1995)
MEDLINE
  95291197
  2 (bases 1 to 4889)
  Jeang,C.L., Liao,T.W., Chiou,S.Y., Kang,P.L. and Shieh,T.Y.
  Direct Submission
  Submitted (25-MAY-1998) Food Science, National Chung Hsing
  University, 250 Kuo Kuang Road, Taichung, Taiwan 402, Republic of
  China
FEATURES
  source
    1. Location/Qualifiers
      /organism="Cytophaga sp. 'Jeang 1995'"
      /strain="Jeang 1995"
      /db_xref="taxon:85173"
      /note="Isolated from soil"
      806..901
      806..3949
      /codon_start=-1
      /transl_table=-11
      /product="S-layer protein precursor"
      /protein_id="AAE21259.1"
      /db_xref="GI:6648627"
      /translation="MAYOPKSYRLKLTATATMVAAGVAPLAFSAASDATGTYKDAV
      VNLIEKGIASGLNKDPFGVOEKIKRGAALMVLTRIGVKGDKADAEFKDINSVAA
      AVAALTYNDIASGATADAFSTGDTITRGOFALYIKENNTIOTSPVITLKGDKDSV
      EYGAADYADAGATVADNKTVEVEKTEKQADAGPVAAVDKVVKGIVTVTSADVAKAN
      KESBACRTVTVKAAOELKIOSVSAINSKQATVFTSAFOKALTSANFDINGLVSDVK
      FSSDRKATVTVSSDFRDOEYSISVGEVDAGAEKAEFGSTFWKAGVVALGS
      TSLGGOTVGLTVKDOAGDKVDKAEVTSYNTLVVTEADDPATVGVKTAANDKLA
      GSTDYYVETTLTPGDSKLANFTYNNVKAESTSISNAGSLEADIDTIGDDMANTAAPFA
      FAPITTSVIGSGSTYELAAFGETNGNPDIYAHVNTGATRTSSPIYATLADNDTAY
      GVVAVTPLSAGTTVTVTKDGRKSPITVATPVOKAIDILNATVTLKSDENEGA
      DAGATFQNGVDKVVKVTSLDQFNKEINKAAQVTVSSSTEGTLTKSADATGTLTD
      OLKAGANELDFAGAKTTNFAILOKDKPVSAATVTVSYFAKADTKRAVTKVSVNT
      DIDPNTLSAIDVVAASEIDANVEVSOAPSIDSSPPVVALDSKGRLEKCTAT
      ADSAADKFEVDNANKLFRKVDNNALTYLRKNTVDVVVEADGVTVLPILAKNTAV
      VPKKAYSTNAVSVKLKSADPNISINDILFGKADAEVLVDCGDIIGVLTARPGSL
      YNRPLVSTIGDKALATGVNVTGGNALITDANFTSDLIETLIGISTTGLAASFTY
      DYVANAANDASTIAGDVTVPVAGKASTFTLVKGIYKDAVLNVEGRAFADASTERA
      AHNLLSPVOLNVSYSK"
      902..3946
      /product="S-layer protein"
      1660 a 838 c 960 g 1430 t 1 others
BASE COUNT
ORIGIN
```

```
alignment_scores:
  Quality: 398.50
  Ratio: 0.855
  Length: 943
  Gaps: 41
```

```
Percent Similarity: 49.417 Percent Identity: 24.178

alignment_block:
us-09-754-947-1 x AF068060 ..
Align seg 1/1 to: AF068060 from: 1 to: 4889

10 Alaasphistprglyle.....Aspsertleasnryrleuvalgluly 24
:::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
905 TGTGATGCTACAGAGATCTACAGAGATGCTGTAACATCTATTATGCAAAA 954
24 sglyalavalysglyasnaspysglymetchpheluproglylysgul 41
|||||  ::::  |||  |||  |||  |||  |||  |||  |||  |||
955 AGGTATCGCTACGCGCTTAACAAAGATACGTTCCGTTCAAGAAAAA 1004
41 eutthrargalaglualalathrmetmetalaaglilleuasnleupro 57
::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1005 TCAACGCTGTCACCGCTGCGATCATGCTACTACCGCTTTAGCGCTA... 1051
58 ileasplyaspsalalysproser..Phealaaspsersinglylnt 73
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1052 TATGGAAGAACGCTAAAGACGCTGCTCAAGATATCAACACAGCCGC... 1099
73 ptyrthrprophellealalavalglulysalagllyallielysgly 89
:::  ::::  |||  |||  |||  |||  |||  |||  |||  |||
1100 .GTACGAAAGCTGTAAACCCACTTAAAGAACCTGGAATCATATACGGTA 1148
90 ..ThrGlyasnlglyphegluproasnlglylysileasparvalsermet 105
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1149 AAACGTCTACTGATTCAGACGACATGCTACATCATCCTGCTGTAATG 1198
122 rproalathrlyspheylsaspleugluThrleuasntrpoglylysgul 139
:::  ::::  |||  |||  |||  |||  |||  |||  |||  |||
123 TAAAGAGTACCTTCAAGATGACCACTGCTATTCAC.....AAAGATG 1286
139 ysalasnilleuvalgluleuglylleaserValgly...ThrGlyasp 154
:::  |||  |||  |||  |||  |||  |||  |||  |||  |||
1287 CGTGAGCTGCTCTTACACTACGACATTCGCAAGCGTGCCTGCTGAT 1336
155 glntprgluprolyslysthrvalthrlysalaglualalaglphel 171
:::  ::::  |||  |||  |||  |||  |||  |||  |||  |||
1337 GCTTTCGGTCAACTGCGACATCACCTGCGGCTCAATTGCTATTTCCT 1386
171 ealalysthrasylys.....GlnpheglyThrglualalala..... 176
:  |||  |||  |||  |||  |||  |||  |||  |||  |||
1387 ATACAAAGCTGAAACATGATTTACCAAAACATCTCCAGTAATCACTTAA 1436
177 .....GlnpheglyThrglualalala..... 183
1437 AAGGTGACAAAGATGATCTGTTGAATAGCGTACGATTATCTGATGCT 1486
183 .....GlnpheglyThrglualalala..... 183
1487 GGTGCAACTGCTGTGACAAACAAGCGGTAAACAGTTGAAGTTAAACAGA 1536
183 .....GlnpheglyThrglualalala..... 183
1537 AGTTAAGATGCTGCTGCTGCCGCTAGTACTGCTGTTGATTAAGTAG 1586
183 .....GlnpheglyThrglualalala..... 183
1587 TTGTGAATATACACAGTTACTTACTCAGCTGTAGATAAAGCTGTAACANA 1636
184 .....LysVal 185
1637 TCTGAAGCATGTCTACTGTAAACAGTTAAAGCTGCTCAAGACTTAAAT 1686
185 lcluseralalysalavalthrThrGlnlysalaglulysphesert 202
:::|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```


697 smllaspseuysalalaleleu..... Asnlelu 708
 3438 AITTTTCTTTTAAACGATTCATTTTGGTAAAGCTGATCGTAGCAACATA 3487
 709 SerValAspGlyValProAlaThrThrAlaValAspAlaThrAlaSerAsnVal 725
 3488 GTTAAAGACGATACGATATTTATCGGTGTCCTTAAACTGGTCCAAACGG 3537
 725 IGluPheValSerAlaAspThrAsnValValAlaGluAsnGlyThrValG 742
 3538 TGGTATCTTA...TACACAAACCCACTGTATCATTTATCTGATCTGATG 3584
 742 IValAlaGlyAlaThrSerIleValVal 751
 3585 GTTAAACATTAGCAACTGGTGTAAACGTA 3613

OM of: US-09-754-947-1 to: EST: * out_format : pfs
Date: Mar 23, 2002 8:28 AM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed_p2n.model -DEV=xlp
-O=/cgr2_1/USPTO.spool/US09754947/rumat_22032002_125444_25431/app_query.fasta_1.854
-DB=EST -QMT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MIMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GCAPOP=4.500
-GCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blomsum2 -TRANS=human40.cdi
-LIST=45 -DOCLIN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=us09754947_rcgcn1_1_4528 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPRY -WAIT -THREADS=1

Search information block:

Query: US-09-754-947-1

Query length: 785

Database sequences: 11351937

Database length: 1077921985

Search time (sec): 1536.890000

Score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_hhc:AK014540	172.00	267.76	9.5e-06	2880	AK014540 Mus musculus 0 day ne
gb_gss:A2538390	149.00	239.70	0.0003	886	A2538390 ENTRO5TR Entameoba h
gb_gss:BH135264	141.00	226.03	0.0020	899	BH135264 ENTNP45TR Entameoba h
gb_hhc:BC003268	141.00	214.68	0.0086	3084	BC003268 Mus musculus, Similar
gb_gss:CN056M7	140.50	223.83	0.0026	1042	AL405221 T7 end of clone A0DAQ
gb_gss:BH135281	138.00	220.81	0.0032	913	BH135281 ENTNOA30TR Entameoba h
gb_hhc:AK004594	138.00	210.19	0.0152	2893	AK004594 Mus musculus adult ma
gb_hhc:BC008524	137.00	204.68	0.0309	4379	BC008524 Mus musculus, Similar
gb_gss:A2538324	135.50	217.04	0.0063	869	A2538324 ENTPO5TR Entameoba h
gb_hhc:AK019969	135.00	206.79	0.0236	2413	AK019969 Mus musculus 8 days c
gb_hhc:AK019450	134.50	205.14	0.0291	2633	AK019450 Mus musculus 14, 17 c
gb_hhc:AK016539	133.00	202.78	0.0394	2382	AK016539 Mus musculus adult ma
gb_gss:A2674287	131.50	210.19	0.0152	877	A2674287 ENTKW3TR Entameoba h
gb_gss:CN0570GA	131.00	209.52	0.0166	943	AL432584 T3 end of clone AZDAQ
gb_hhc:BC008697	131.00	201.49	0.0465	2056	BC008697 Homo sapiens, clone I
gb_hhc:AK004893	129.50	195.53	0.0998	2983	AK004893 Mus musculus adult ma
gb_hhc:AK004966	129.00	195.06	0.1061	2864	AK004966 Mus musculus adult ma
gb_gss:CN05709U	128.50	204.72	0.0307	915	AL423352 T3 end of clone AZDAQ
gb_hhc:AK013048	128.50	197.82	0.0744	1936	AK013048 Mus musculus 10, 11 c
gb_hhc:BC011416	128.50	196.53	0.0878	2226	BC011416 Homo sapiens, Similar
gb_hhc:AK004862	127.50	192.11	0.1547	2993	AK004862 Mus musculus adult ma
gb_hhc:AK016519	127.00	188.19	0.2558	4581	AK016519 Mus musculus adult ma
gb_hhc:BC006984	127.00	197.05	0.0832	1598	BC006984 Homo sapiens, Similar
gb_hhc:AK017839	127.00	196.50	0.0861	1695	AK017839 Mus musculus 8 days g
gb_gss:CN056ZQT	125.50	199.17	0.0626	963	AL422667 T3 end of clone AZDAQ
gb_hhc:AK014398	125.00	199.13	0.1995	2345	AK014398 Mus musculus adult ma
gb_hhc:BC011226	124.50	198.40	0.2492	2582	BC011226 Mus musculus, Similar
gb_gss:CN0560JDE	122.50	194.04	0.1208	969	AL410520 T3 end of clone AZDAQ
gb_gss:CN056P84	121.50	191.06	0.1705	1114	AL409978 T3 end of clone XAVDA
gb_hhc:AK014839	121.50	180.24	0.7095	3609	AK014839 Mus musculus adult ma
gb_gss:CN0570US	120.00	189.44	0.2180	1009	AL424106 T7 end of clone AZDAQ
gb_hhc:AK019472	119.50	178.29	0.9106	3087	AK019472 Mus musculus 0 day ne
gb_hhc:BC013952	119.00	178.08	0.9358	2882	AK013952 Mus musculus 13 days
gb_gss:CN0570DL	118.00	186.93	0.3006	917	AL423487 T7 end of clone AZDAQ
gb_hhc:AK015000	118.00	180.65	0.6693	1806	AK015000 Mus musculus adult ma
gb_hhc:BC011801	118.00	179.25	0.8057	2113	BC011801 Homo sapiens, clone I
gb_hhc:BC003664	118.00	177.88	0.9601	2451	BC003664 Homo sapiens, Similar
gb_est1:AA264459	117.50	188.87	0.2345	678	AA264459 LD08971.Sprline LD Dros
gb_hhc:BC003279	117.50	188.55	0.8805	2078	BC003279 Mus musculus, Similar
gb_est2:BG638449	117.00	186.53	0.3164	797	BG638449 LD25659.Sprline LD Dros
gb_hhc:AK004574	117.00	174.42	1.50	2971	AK004574 Mus musculus adult ma

gb_gss:A2530606 + 116.50 185.18 0.3763 882 I A2530606 ENTCA11TF Entameoba
gb_gss:A2691742 - 116.50 184.49 0.4114 908 I A2691742 ENTKW46TR Entameoba
gb_gss:A2683303 + 116.00 183.59 0.4615 913 I A2683303 ENTKW21TR Entameoba
gb_hhc:AK011815 + 116.00 178.53 0.8828 1581 I AK011815 Mus musculus 10 da

seq_name: gb_hhc:AK014540

seq_documentation_block:

LOCUS AK014540 2880 bp mRNA HTC 05-JUL-2001
DEFINITION Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4631429H07, full insert sequence.

ACCESSION AK014540
VERSION AK014540.1 GI:12852460
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library
clone:4631429H07.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

1 (bases 1 to 2880)
Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning
Methods in enzymology. 303, 19-44 (1999)
99279253
PUBMED
10349636

2 (bases 1 to 2880)
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome research. 10 (10), 1617-1630 (2000)
11042159

3 (bases 1 to 2880)
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system - 384 format
Sequencing pipeline with 384 multicapillary sequencer
Genome research. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861

4 (bases 1 to 2880)
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 (bases 1 to 2880)

Adechi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arikawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Futuno, M.,
Hara, K., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,
Iomtani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,
Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
Miyazaki, A., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Okazaki, Y.,
Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H.,
Sasaki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A.,
Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
url: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

COMMENT


```

412 luserleuasntrhgluvalala.....Val 420
      ::::::::::::::::::::
1818 CTGAGCTTAAAGACAGATAGAGAGCTAGACTAGATTACAGACAGAA 1867
421 ValAspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAlaPr 437
      ::::::::::::::::::::
1868 ATGGAAGTTTACAGAGTAA.....1888
437 ovalysvalthrvalysasperlysglyalaleuvalserlstr 454
      ::::::::::::::::::::
1889 .....CAGACCTCGAGCGCTCGCT.....CATG 1913
454 hrvalgluilegualaphleaglinsalawetlysasplilelys... 469
      ::::::::::::::::::::
1914 CTAAAGATGAGACCATCGAGCCCAAGCTGATCAAGATCAATTAAGAG 1963
470 .....leuglyusthrasnvalalaleuSerThrlyaspva 482
      ::::::::::::::::::::
1964 AAGCAGACAGCTCGAGCGCGTCAAGCAGAGCTGAGAGTGCAGAGA 2013
482 lthrAspLeuLysValLysAlaProValLeuAspGlnTyrGlySGluP 499
      ::::::::::::::::::::
2014 CCAGCACCTA.....GTGGAATGAGAGACACCTTAACAACTGC 2054
499 heThrAlaProValThrValLysValLeuAspLysAspGlyLysGluLeu 515
      ::::::::::::::::::::
2055 AGCAGCGCGAGATTAAAGTAAAGAGCTA.....2083
516 LysGluGlnLysLeuGluAlaLysTyrValAlaSnArgLysLeuValLeuAs 532
      ::::::::::::::::::::
2084 .....GAGTTCTGAGGCCAAGTACAGTGAACAAAGCGAGGTCAATT.. 2125
532 nalAlaGlnGlnLysAlaGlyAsnTyrThrValValLeuThrAla.... 547
      ::::::::::::::::::::
2126 .....GCTATTTTACATCGCAGCTCGCTGCTGCA 2156
548 .....LysSerGlyGluLysGluAla 554
      ::::::::::::::::::::
2157 AAGAAAAGCTCTTGATCTGATGCGCTCGAAAGCCAAATTCGGAAGGT 2206
555 LysAlaThrLeu...AlaLeuGluLeuLysAlaProGlyAlaPheSerly 570
      ::::::::::::::::::::
2207 AAAGCTGAGCTCGAGACACTTACACAGCAGCTTGAAGGAGCT.....GA 2250
570 sphleuValArgLysLeuAspThrGluLeuAsp.....581
      ::::::::::::::::::::
2251 AAGCAGATTAAAACTTAGAGCTGAAGAGCAGCTGAAGTAGCAAG 2300
582 ..LysTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerVal 597
      ::::::::::::::::::::
2301 CTATATGACATTACCAAGAGCTCCAGAGAAAGAGCTAGTCTCAGTGGT 2350
598 LeuProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaAlaG1 614
      ::::::::::::::::::::
2351 CTGCAG...CAGACTTGAATCAAGTCAATCAAGTGAAGGAGCTTGGGA 2397
614 uLeuLysValThrThrThrAsnLysGlnGlyLysGluValAspAlaThrA 631
      ::::::::::::::::::::
2398 GAAGAAGACTTCAGACTTTGAAGAAA.....AAGTTTCCAGTACTTCA 2441
631 spAlaGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGly 647
      ::::::::::::::::::::
2442 AGGAGCGGCTCTGCTCAGACCAAGAGTGCAGAGATACGTAATAAATG 2491
648 AlAlaLysAlaGlyGluThrTyrLysVal 656
      ::::::::::::::::::::
2492 CACCAAAAGCAGCAACAGTTAACGTG 2518
seq_name: gb_gss:AZ538390
seq_documentation block: 886 bp DNA GSS 14-NOV-2000
LOCUS AZ538390
DEFINITION ENTB0051F Entamoeba histolytica Sheared DNA Entamoeba histolytica

```

```

ACCESSION genomic, DNA sequence.
AZ538390
VERSION AZ538390.1 GI:11443284
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 886)
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: enta@ligr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 28
High quality sequence stop: 744.
Location/Qualifiers
1..886
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/notes="Vector: PHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
light size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1999)."
```

```

BASE COUNT      421 a      111 c      194 g      160 t
ORIGIN
alignment_scores:
Quality: 149.00      Length: 406
Ratio: 0.776      Gaps: 17
Percent Similarity: 47.291      Percent Identity: 24.138

```

alignment block:

us-09-754-947-1 x AZ538390 ..

```

Align seg 1/1 to: AZ538390 from: 1 to: 886
114 LysLeuAspThrLysValAsnGlyThrProAlaThrLysPheLysAspLe 130
      ::::::::::::::::::::
29 AAGAAAGAGCTTAAGACAGAGTGAAGAGAGCAGAGAAAG...AAGGCTGC 75
130 uGluThrLeuAsnTrpGlyLysGluLysAlaAsnLysLeuValGluLeuG 147
      ::::::::::::::::::::
76 TAAAGTACTTAAGAAAGCAGAAAGAGCAGCT.....109
147 LylIeSerValGlyThrGlyAspGlnTrpGluProLysLysThrValThr 163
      ::::::::::::::::::::
110 .....GAGCTTAAGAGAGCTGCTACT 130
164 LysAlaGluAlaAlaGlnPheIleAlaLysThrAspLysGlnPheGlyTh 180
      ::::::::::::::::::::
131 AATATCA...GCAGCTAAAGCTACTGAAAGGCTAAGAAATCTATT..... 172

```

```

180  rGUAAlaLalysValGluSerAlaLalValThrGlnLysValG 197
    ||| ||||| ||||| ||||| ||||| ||||| |||||
173  .GAAGAAGCTAAATCAGATCAAGAAGGCA.....GCTAAGAAGTAG 215
197  LuValLysPheSerLysAlaValGluLysLeuThrLysGluAspIleLys 213
    || ||||| ||||| ||||| ||||| ||||| |||||
216  AAAGAATAATG...AAAGAATCAAGAATCATTTGAAGAGTAAAGATTAAAG 262
214  ValThrAsnLysAlaAsnAsnAspLysValLeuValLysGluValThrLe 230
    ||| ||||| ||||| ||||| ||||| ||||| |||||
263  GTTCACATCAAGACT.....GCTAAGAAGAGCTGCTAA 294
230  uSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsnLeuAla 247
    :||| ||| ||||| ||||| ||||| ||||| |||||
295  AGCAGAAAGAGAGCTGAAAGCAATTAAGCAAGATTAAAG..... 337
247  LalysGlnThrTyrThrValAspValAsnLysValGlyLysThrGluVal 263
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
338  .....GATGTTAAGAAGAGCACTAAGACGCTGCT 367
264  AlaValGlySerLeuGluAlaLysThrLeuMetAlaAspGlnThrVal 280
    :||| ||||| ||||| ||||| ||||| ||||| |||||
368  CAACTCAAGAAGTACAACT..... 388
280  lValAlaAspGluProThrAlaLeuGlnPheThrValLysAspGluAsnG 297
    ||| ||||| ||||| ||||| ||||| ||||| |||||
389  .....CAAGTTAAGAAGAGCAAGCAAGATG 413
297  LyrThrGluValLysSerProGluGlyLeuGluPheValThrProAlaAla 313
    :||| ||| ||||| ||||| ||||| ||||| |||||
414  CTAAGAAAGCTCTTTAAAGAAATCA...AAATTAAGAAATCAATTAGCT 460
314  GluLysIleAsnAlaLysGlyGluIleThrLeuAlaLysGlyThrSerTh 330
    :||| ||||| ||||| ||||| ||||| ||||| |||||
461  AAGAAGATTAGTAAG.....ACTTCTAA 483
330  rThrValLysAlaValTyrLysLysAspGlyLysValAlaAlaGluSerL 347
    :||| ||||| ||||| ||||| ||||| ||||| |||||
484  GAATGCTAAAGCTATTGCCAAGAGCTAAGCAAGAGCTGCTAAAGAGG 533
347  ysgLysValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsn 363
    ||| ||||| ||||| ||||| ||||| ||||| |||||
534  CTGAAGAAGAA.....GCAGAGAGAGAAAGCTATTAGCAAGCAAGAGCT 577
364  TrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLys 380
    :||| ||||| ||||| ||||| ||||| ||||| |||||
578  TACACT.....AAGAAGATACCAAGAGATTGCTAAGAGATTAAGAA 621
380  sGlnAsnAsnLysValTyrGluGlyAsnAsnAlaTyrValGlnValGluL 397
    :||| ||||| ||||| ||||| ||||| ||||| |||||
622  AAAACCTTAACAGGTT..... 637
397  euLysAspGlnPheAsnAlaValThrThrGlyLysValGluTyrGluSer 413
    :||| ||||| ||||| ||||| ||||| ||||| |||||
638  .....CAATCTAAG 646
414  LeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLysValThrVal 430
    :||| ||||| ||||| ||||| ||||| ||||| |||||
647  TTGAAGAAAGAGCTTAAGAAAGTAAAGAAAG..... 676
430  lleuSerAlaGlyLysAlaProValLysValThrValLysAspSerLysG 447
    :||| ||||| ||||| ||||| ||||| ||||| |||||
677  ....AATGATAAAGAGGCTTGTCTAAGATAGCTAAGAAAGACAAAGAAAG 722
447  lLysAlaLeuValSerHisThrValGluIleGluAlaPheAlaGlyLys 463
    :||| ||||| ||||| ||||| ||||| ||||| |||||
723  CACTTGTCTAAATTAATCT.....CAGAAA 745
464  AlMetLysAspIleLysLeuGlyLysThrAsnValAlaLeuSerThrLys 480
    :||| ||||| ||||| ||||| ||||| ||||| |||||
746  GTTAAGAAAGCTCAAGAAAGATAGAAAGAGTATTGCTAAATTAAGCTAA 795

```

```

480  sAspValThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyL 497
    ||| ||||| ||||| ||||| ||||| ||||| |||||
796  AAAGACACGAAGGCTTGTCTAAATTAATGCTAAGAAAGAGAG..... 838
497  ysgLupheThrAlaProValThrValLysValLeuAspLysAspGlyLys 513
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
839  .....AAAGCTATTGCTAAATTAAGCTAAG 862
514  GluLeuLysGluGlnLys 519
    :||| ||||| ||||| ||||| ||||| ||||| |||||
863  AAGGCTAAGAGACCAAA 880
seq_name: gb_gss:BH135264
seq_documentation_block: 899 bp DNA GSS 07-AUG-2001
LOCUS BH135264 Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION BH135264
VERSION BH135264.1 GI:15094325
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 899)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
CONTACT: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: entae@ligr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 801.
FEATURES
source
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site1.Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999)."
```

alignment_scores:

Quality:	141.00	Length:	309
Ratio:	0.839	Gaps:	15
Percent Similarity:	54.369	Percent Identity:	26.214

alignment_block:

US-09-754-947-1 x BH135264/rev ..

1221 AAGATGAAGACCTTCATATG.....GCTGAGAAAGACTAAACT 1258
142 eluValgluleuglyIleSerValglYthrGlyaspGlntrpGluProL 159
:::|||||:::|||||:::|||||
1259 TATGACTGACTGCGGTGATGCAAGAACTTAATTCAGACGCTTGAAACAG 1308
159 yslYthrValThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAsp 175
|||:::|||||:::|||||:::|||||
1309 ATAAAGGGATGCTAATTAACCGAGAGCAAGCGCAAAATGCTTGAGACCTG 1358
176 LysGlnPheGlyThrGluAlaAlaLysValGluSer...AlaLysAlaLys 191
:::|||||:::|||||:::|||||:::|||||
1359 GAACGAAAGAAAGATGAATGCTGCTGACCTGCTACTCATATCAAAACACAT 1408
191 LThrThrGlnLysValGluValLysPheSerLys..... 202
:::|||||:::|||||:::|||||
1409 GACTTAACCCAGGAGAGCAATGCGGGAGACAGAAAGAAAGTGTGAAAGAG 1458
203 ..AlaValGluLysLeuThrLys.....Glu 210
|||:::|||||:::|||||
1459 CTGCTTTTGGAGAACTTGAAAAAGCCTTGAGTACAGCCCAAAACTGAA 1508
211 AspIleLysValThrAsnLysAlaAsnAsnAspLysValLeuValLysGlu 227
|||:::|||||:::|||||:::|||||
1509 GATGCACAGACAGAAATGAATGATGAGATGATGATGAACAGATGAAGCTGT 1558
227 uValThrLeuSerGluAspLysArgSerAlaThrValGluLeuThrYserA 244
|||:::|||||:::|||||
1559 TGAGAGAGCCGAGGAGGAGGAGCC.....CTCCGCTTGAGCATG 1599
244 snLeuAlaAlaLysGlnThrYThrValLysValAsnLysValGlyLys 260
:::|||||:::|||||:::|||||
1600 AGCTGAGCGGCTGAGGAGGAGGCTGCCACATGCAAAAGAACTCAGAA 1649
261 ThrGluValAlaVal..... 265
:::|||||:::|||||
1550 GAACAAGTGTGTCACATCCAGAACGTCATGCAGAGAGCTGCCAGCAA 1699
266GlySerLeuGluAlaLysThrIleGluMetAla 277
|||:::|||||:::|||||
1700 AGACGAGAGCTGAGCAGAGAGCTTGAGCCCGGAAAGGAGCTGCGAG 1749
277 spGlnThrValAlaLysPglupProThrAlaLeuGlnPhe.....Thr 291
:::|||||:::|||||
1750 AGCAGATGAGAAATAGCTCTCGAAAAGAGTCGATAGAAATATTGAAAGCTC 1799
292 ValLysAspGluAsnGlyThrGluValLysSerProGluGlyIleGlu.. 307
:::|||||:::|||||:::|||||
1800 ACCCAAGAAAAAGAGCAGCAAGAAATCTTGCTTGAGAAAGATTAGAGCT 1849
308PheValThrProAlaAlaGluLysIleAsnAlaLysG 320
|||:::|||||:::|||||
1850 GCAGAAAAAACCCATCTTACAGAGAGTGAATAAATCTCCAGAACACTG 1899
320 Ly.....GluIleThr 323
|||:::|||||
1900 GGCAGAAAGCAGAGGCTTACCGAACCAAGATTCCTTGGAATTGGAACCTCT 1949
324 LeuAlaLysGly.....ThrSerThrValLysAlaLys 335
|||:::|||||:::|||||
1950 TTGGAAAAAAGATTACAAAGAAACACAGTCAGAGCATCTGGCGT 1999
335 LTrp.....LysLysAspGlyLysValAlaLysGlu 346
|||:::|||||:::|||||
2000 TCATCTGGAAGCTGAGAGAAATTAAGCAATTAAGAACTCACACCTCTGG 2049
346 erLysGluValLysValSerAlaGluGluAlaAlaValAlaSerIleSer 362
:::|||||:::|||||
2050 CTGAGCAGCATAGACAGAAAGTGGAGGCTCCAGCAGCAGCAGCAGCAGC 2099
363 AsnTrpThr.....ValAlaGluLysAsnLysAla..... 372
|||:::|||||:::|||||
2100 CTGTGAGACTGAGAGACTCAAAAGCCTCTCTCAGAGCATCAGCTGCTGT 2149

373 ...AspPheThrSerLysAspPheLysGlnAsnAsnLysValTrpGluG 388
:::|||||:::|||||
2150 GGAGAGCTCAGAGAGAGATATCACAGAGAGAAAGATGCACACTAGCAGAG 2199
388 LysAsnAlaLysValGlnValGluLeuLysAspGlnPheAsnAlaVal 404
:::|||||:::|||||
2200 AGAAGAGAGTCTCTTCCAGGCCCATACAGAAC...ATGAATGAAAG 2246
405 ThrThrGlyLys.....ValGluTrpGluSerLeuAsnThr 416
|||:::|||||:::|||||
2247 ACCTTGGAAGAACTTGACACAGACAAATGCACACTGGAATCTGTCTTTC 2296
416 rGluValAlaValVal.....AspLysAlaThrGlyLysValT 429
|||:::|||||:::|||||
2297 TGAGCTGTCAAGACACATGAGAGCTCGGACACAGCTTCAGAGAGCTTT 2346
429 hrValLeuSerAlaGlyLysAlaProValLysValThrValLysAspSer 445
:::|||||:::|||||
2347 CTGTCTTAAGAGTGATGCAGATAAATGAACAGAGCTTGAGGCCGAG 2396
446 LysGlyLysAlaLeuValSerHisThrValGluIleGluAlaPhe.... 460
|||:::|||||
2397 CTGAGAGAGCAGCGCGCTCACACAGCGTGAAGCTTGCCACATCAGTGA 2446
461AlaGlnLysAlaMetLysAspIleL 469
|||:::|||||
2447 GCAACAAGAACTAACCCGCCAGAGCTGAGAGAGGCACTGAAGAAT...G 2493
469 yslLeuGluLysThrAsnValAlaLeuSerThrLysAspValThrAspLeu 485
:::|||||:::|||||
2494 AGCTCAGTCGCTGGGCTCTGCTGACGAGAGAGGAGGAGCACCCTGAGA 2543
486 LysValLysAlaProValLeuAspGlnTrpGlyLysGluPheThrAlaPr 502
:::|||||
2544 GAGCGGCGAGCGCGGCTCAGAGACTGGAACCTCAT..... 2579
502 oValThrValLysValLeuAspLysAspGlyLysGluLeuLysGluGlnL 519
|||:::|||||
2580CTTCAGAAAGTCTGCCGGGAGCTCCAGCAGCGCT 2613
519 yslLeuGluAlaLysThrValAsnArgLysLeuValLeuAsnAlaLysGly 535
|||:::|||||
2614 TGCCCAAACTAGATCTCTTGACCTCTGAG.....CAGAGCGCCGACAGC 2657
536 GlnGluAlaGlyAsnTrpThrValValLeuThrAlaLysSerGlyGluLys 552
:::|||||
2658 GAGCAGGCGAGGTGGTATGAGAGCAGCTG...GCCCAATGACAGCAAAA 2704
552 sGluAlaLysAlaThrLeuAlaLeuGluLysAlaProGlyAlaPheS 569
|||:::|||||
2705 G.....GTGTTAGACTGTGAAACAGAGAGAGCCCTTC 2736
569 erLysPheGluValArgLysLeuAspThrGluLeuAspLysValThr 585
:::|||||
2737 TTACCAAGCAGAGTAGTGAGATGCAAAACACAC...AAAAACATGTGTGT 2783
586 GluLysGlnLysAsnAlaMetThrValSerValLeuProValAspAl 602
|||:::|||||
2784 GAGAAATTAGATGCTCAGAGAGCTCAAGTGCAGCAGCTGGAGAGACAG 2833
602 aAsnGlyLeu.....ValLeuLysGlyAlaGluAlaLysG 614
:::|||||
2834 GAGTGAGCTGTGAGAGAGAAAGTCAAGTGCCTAGCCACCTCCAGAGCTCAC 2883
614 LuLeuLysValThrThrAsnLysGlu 623
:::|||||
2884 AGCTCAAGAACAGTACTGTGGAGAAAGAG 2912

seq_name: gb_gss:CNS06MA7

seq_documentation_block: 1042 bp DNA GSS 17-JUN-2001
LOCUS CNS06MA7

[illegible]

```
|||||.....|||||.....
876 .....GACGCCGAAAAATCAAGACCAACGCC... 905
522 AAlaLysTrValAsnArgLLeuValLeuAsnAlaIaGlyGluAl 538
906 .....|||||.....
538 aGlyAsnTrThrValValLeuThrAlaLysSerGlyLysGluAlaL 555
923 G.....GAAGTAAAAAG 936
555 ySaLaThrLeuAlaLeuGluLeuLysAlaProGlyAla 567
937 CTGCTACTTTAGCTTTAGACAAACAAGCTGCTGCC 974
seq_name: gb_gss:BH135281
```

```
seq_documentation_block: 913 bp DNA GSS 07-AUG-2001
LOCUS BH135281
DEFINITION ENTOM30TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION BH135281
VERSION BH135281.1 GI:15094342
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 913)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: entae@ligr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 13
High quality sequence stop: 754.
Location/Qualifiers
1. 913
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/Note="Vector: pHO51; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Bartell, Oxford University Press, 1999)."
```

```
BASE COUNT 439 a 116 c 162 g 196 t
ORIGIN
alignment_scores:
Quality: 138.00 Length: 360
Ratio: 0.802 Gaps: 14
Percent Similarity: 47.778 Percent Identity: 22.778
```

```
alignment_block:
US-09-754-947-1 x BH135281
Align seg 1/1 to: BH135281 from: 1 to: 913
```

```
171 lLeaLaLysTrAspLysGlnPheGlyThrGluAlaAlaLysValGluSe 187
:::|||||:::|||||:::
10 CTAAAGAAAGACAAAGACACTTCTCTAAATTTACTAAGAAAGTT..... 54
187 rAlaLysAlaValThrThrGlnLysValGluValLysPheSerLysAlaV 204
:::|||||::: ||| :::|||||:::
55 .....AAGAAAGTTCAAAAGAAAGAAATTAAGAAAGCTA 85
204 aGluLysLeuThrLysGluAspLysValThrAsnLys...AlaAsn 219
:::|||||:::|||||::: ||| |||:::
86 TTGCTAAATTTAGCTAAAGACAAAGAGCGCTCTGCTAAATTTCTTAAG 135
220 AsnAspLysValLeuValLysGluValThrLeuSerGluAspLysArgSe 236
:::|||||::: ::| :::|||||:::
136 AAGCAACAAGAAAGCTATTGCTAAATTTAGCT.....AAGCAAGC 173
236 rAlaThrValGluLeuThrLysSerAsnLeuAlaAlaLysGlnThrTrp 253
:::|||||:::..... |||||||:::
174 TAAAGACGCCAAAGAAAGTTGCTAAGACTGCTGCTAAAGAGCTCAAAAG 222
253 aLAspValAsnLysValGlyLysThrGluValAlaValGlySerLeuGlu 269
:::|||||:::|||||::: |||:::
223 ..AAGCAACTAAGCTGCTGCAAGAAAGAGAGCTTTAAAGCTTAATTTCT 270
270 AAlaLysThrLLeuGluMetAlaAspGlnThrValValAlaAspGluPro 286
|||||
271 GCTTAA..... 276
286 rAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValAlaSerP 303
276 ..... 276
303 roGluGlyLLeuGluPheValThrProAlaAlaGluLysIleAsnAlaLys 319
277 .....AAG 279
320 GlyGluLeuThrLeuAlaLysGlyThrSerThrThrValLysAlaValTy 336
|||||
280 GCTGAAGAAAGAACTGCTTAA.....AAGGTTAAAGCGTTAA 317
336 rLysLysAspGlyLysValValAlaGluSerLysGluValLysValSera 353
|||||:::|||||::: :::|||||:::
318 GAAAAATGAATCTAAATTAATCAAAAGAAATTAATTAAG...AAACTGCTA 364
353 lAgLuGlyAlaAlaValAlaSerIleSerAsnTrpThrValAlaGluGln 369
:::|||||::: |||
365 AGAAAGCGCCAAAGCTGCTAAG.....AAGAAAGCT 396
370 AsnLysAlaAspPheThrSerLysAspPheLysGluAsnLysValTy 386
397 TTAAGAGCTTAATTAATTCGTTAAAG..... 423
386 rGluGlyAspAsnAlaTrpValGluValGluLeuLysAspGlnPheAsnA 403
|||||:::|||||::: |||
424 .....GCTGAAGAAAGAACTGCTTAAATTAAG 454
403 lValAlaThrThrGlyLysValGluThrGluSerLeuAsnThrGluValAla 419
|||||
455 CAGTTTAA.....AAGAAAGAGCTTAA 477
420 ValValAspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysAl 436
:::|||||:::|||||::: ||| |||
478 TTAATCAAAAGAGTAATTAAGAACTTCAAAAGCTCTT...AACAAAGC 524
436 aProValLysValThrValLysAspSerLysGlyLysAlaLeuValSerH 453
| ||||||| |||:::|||||
525 TGCTAGTAAAGTAACCAAGAGAGTACTTAA.....A 556
```


131 luthleuasntpcllysgluysalaasnlle.leuvalgluleucl 147
||||| : : : : :
409 AAGCACTACACACACGATCCGAGCCTCATCTGCTG..... 450
147 yileserValglYthrGlyaspGlnTrp.....GluProL 159
: : : : :
451 .CTCAGCTGCACACCTGAGCGCTCTGCTACCAAGTGTGACGAGAG 499
yslysthrValthrlysalaglualalaglInpheIlealalystrasp 175
||| : : : : :
500 TCAAGTACTGTAAGCTCAACCCGATGGGCCAAGTGATTATGTTTGA 549
176 lysglInpheglYthrGluAlalalysValgluseralalysalavalThr 192
||| : : : : :
550 AAACAGCTGCCGTAAAGACTTCAAAACCACACAGAAA.....AA 590
192 rThcGlnlysalgluVallyspheSerlysalavalglulysleuThrL 209
: : : : :
591 TAAAGCACACTGAGCTCGAA...AATTAATAATTGAGAAAGAGACTA 637
209 ysgluaspIlelysalThrAsnlysalAsnAsnAspLyvalleuVal 225
||| : : : : :
638 AAATATGAA...CAAGAGACAGAAATCGAAAC.....CTGGCT 675
226 lysgluValthrLeuSerGluAspLysArgSerAlathrValgluleuTy 242
||||| : : : : :
676 AAAGAACTATTCCCATGACTCTCTCCAGATACGTGAAAGGACT 725
242 rSerAsnleuAlalalysglnThrlyThrValaspValasnllysalG 259
||||| : : : : :
726 CAGTAATTTGGGGAATCTGTTTCTTCAATGCAATGATGCAAGACTTGT 775
259 lylysthrGluValAlalalyserleuGluAlalythrIleGlnmet 275
: : : : :
776 CACAAAGCCAGTGCTTACAGACTACTA.....AAGAGTGAAGATG 819
276 AlaspGlnthrValAlalaspGlnProthrAlaleuGlnpheThrVa 292
: : : : :
820 TCTGGA...ACGATTGTAAATAAGAGCCACTCATCTGGCCTAACCA.. 864
292 llyaspGlnAsnGlyThrGluValValserProGluGlyIleGluPheV 309
864 864
309 alThrProAlalaglLyIleAsnAlalygluIleThrleuAla 325
||| : : : : :
865 .GAACCTTTAGAACTAAACCTCGAGCTCCTTACTTATGTTACC 912
326 lysglYthrSerThrValAlalyValTyrlYsLysaspGlyLysVa 342
913 ATG.....AGCCAGTT 923
342 lvalalagluserlysgluVallysalserAlaglulalalalava 359
: : : : :
924 TCTCAGTGAAGTGCACAGAACAAA..... 948
359 laseIleaserAsnTrpThrValalaglInAsnlysalaspPheThr 375
949AAGCGAGTTGTGACA 963
376 serlyasp...PheLySGlnAsnAsnlysalYrGluGlyaspAsnAl 391
964 CTTAAAGAGCTTTTCTTCTCAGGCTGTAAAAAACCAACACGCTTTTAAAG 1013
391 aTyRValGlInValGluLeuLysaspGlnPheAsnAlalValThrThrGlyL 408
: : : : :
1014 GTACACACACACAGACAGCCAGAGCTGCTTCCGTACTACTGATGGGA 1063
408 ysValgluTyRgluSerleuasnthrGluValAlalValaspLysAla 424
||| : : : : :
1064 TGACAGCGCAAAACCAAGAGCTGAGTAAAGCAATTCTT...AAAGCA 1110

425 ThrGlyLysValThrValleuSerAlaglyLysAlaProValLysValThr 441
||||| : : : : :
1111 TTGGTAATTTCTACT.....GAAAAATTGCATGAAGAAGTAAAAAATA 1154
441 rValLysasp...SerLySGlyLysAlaleuValSerHsthThrValGluT 457
||||| : : : : :
1155 AGTTAAAGATTATGAAGAAAGCAATCCCGAGTTTGTGGACCCGA 1204
457 lgluAlalpheAlaglInlysalameLysAspIleLysleuGluLys... 472
1205 TC.....TTGGTGGCAGCTGACTAGTACGATATGTGTCATATGC 1248
473 ThrAsnValAlaleuSerThrLysaspValThrAspLeuLysValLysAl 489
1249 AGGACTGCTCTCTTAGTCATGAATCGTTCTTATTTGTCCTT..... 1293
489 aproValleu...AspGlnTyRglYlysgluPheThrAlaproValThrV 505
||||| : : : : :
1294 .CCAGTTTATGATGATCAGAGTGTAAAGAAAATTAATGATAAAATG 1342
505 allysVal...LeuasplysaspGlyLysgluleuLysgluInlystreu 520
||||| : : : : :
1343 TCAAAATGACATGAGAGAGAGATTAAGACAGTGCAGAGCAAGAAA... 1389
521 GluAlalyTyRValAsnArgGluLeuValleuAsnAlalaglyGlnG 537
: : : : :
1390 GATGACACTACATGAATAATCAAGAGCGATCTCCGTACGAGG..... 1431
537 uAlagLysnthrThrValValleuThrAlalyserGlyGluysGluA 554
1432ACAAAGCAAGCACTTACAGAAAAGG 1456
554 lAlysAlaThrleuAlaleuGluLeuLysAlaProGlyAlaPheSerLys 570
1457 CAAGAAACAGAGCCAAAAGCAGCCCAAGAACCAAGCAAGCAAGCAAAA 1506
571 pheGluValArgGlyLeuasp...ThrGluLeu.....AspLysTyRva 584
1507 ATTCAAGAAATTTCTTCACTTCATAGCTCTGCCACTGACTAC.. 1554
584 lThrGluGlnAsnGlnLysAsnAlameThrValserValleuProVala 601
1555 .ACGGAAGACATGACCTGAAGCTGACACA..... 1584
601 spAlaAsnGlyLeuValleuLysgluAlagluAlalagluleuLysVal 617
1585GCACCTGGCGGAGAGTGAAGTG 1608
618 ThrThrThrAsnLysgluGlyLysgluValaspAlathrAspAlaGlnVa 634
||| : : : : :
1609 GATACCGACTCCACCACATGCTTCACAGAGAGGCCACACAGATAGACT 1658
634 lThrValGlnAsnAsnserValIleThrValglYglnGlyAlalySalag 651
1659 GTCTGTATACCAAGAAGAT..... 1677
651 lygluThrTyRlysalThrValValleuAspGlyLysleuIleThrThr 667
1678TTGATGGCCCAAGAGAGCATGATA 1701
668 HisSerPheLysValValaspThrAlarProthrAlalySGlyleuAlaVa 684
1702 GAAAGGACACCTGATGTGCAGAAAGCCCAAGAGACCTAGAGTGAAG 1751
684 lGluPheThrSerThrSerleuLysgluValAlaProAsnAlaspLeuL 701
1752 TGCTAACACCCAGAGTATCTGGGAGATTGACGCT..... 1788
701 ysAlaAlaleu.LeuasnlleLeuSerValaspGlyValProAla 715
1789GCTCCTGAATGTCCTAGGATTTCAATGTGCTTCC 1825
seq_name: gb_hlc:BC008524

1130 CACCCTCCGACGATCTCTGAGACTGTGATTAATTAGAGCGCAACTGAA 1179
338 LysAspGlyLysValAlaGluSerLysGluValLysValSerAlaGlu 354
1180 AAGTTGGGACAGCGCTGACT.....GAGCTATGCCAAGGCCAA 1220
354 uGlyAlaIaValAlaSerIleSer.....AsnTrpThr..... 365
1221 CAGTTTACACTCTCCAGTCTGTGCCCCCTTCGCTTCATCGTTTCA 1270
366ValAlaGluIaAsnLysAlaAspPheThrSerLysAspPhe 379
1271 TCATTGTGAACAAGGCGAAGCCTGCGCAAAATCTCAGCAAAATGCCA 1320
380 LysGluAsnAsnLysValTrpGluLysAsn..... 390
1321 AAGTTTCACATTGAGTTCCAGAAAGTGAAGCAAGATCACCCTGAGGG 1370
391AlaTrpValGluValGlu.....LeuL 398
1371 CCCCACAGAGCATGTGAATGTAGCCCAAGAACAGATTGAAGAAATGCTCA 1420
398 LysAspGluPheAsnAlaValThrThrGlyLysValGluTrpGluSerLeu 414
1421 AAGACCTATTAAAT.....AGGATGGACTATGTAGAGATC 1455
415 AsnThrGluValAlaValAlaAspLysAlaThrGlyLysValThrVal.. 430
1456 AACATTGACCTAAATTCACAGACACTCATTTGAAAGACGCGAGCCAA 1505
431 .LeuSerIaGlyLysAlaProValLysValThrValLys.....A 444
1506 CATCAACAGAAATCAAAAGACCCAGTCAAGTGTGTGCGCATCCCCCTG 1555
444 sPseLysGlyLysAlaLeuValSerIleThrValGluIleGluAlaPhe 460
1556 ACAGTGAA.....AAGAGTAACTGATCCGAAATTGAGGGAAC 1593
461 AlaGlu..... 462
1594 CCACAGGGGTTCAGACAGCGGAGGAGCTGTGAGCTTGTCTCG 1643
463LysAlaMetLysAspIleLysLeuGlu..... 471
1644 CATGGAATAATGAGCGCACCAAGATTAATCATGAGCAAAAGATTTCATC 1693
471 471
1694 GCACAATCATTTGGCGAAGAAAGTGAACGATTTCGAATAATTCTGACAAA 1743
471 471
1744 TTCCAGAGCTTATCATCACTTCCAGACCCACACAAAAGATGATAT 1793
472LysThrAsnValAlaLeuSerThr..... 479
1794 TGTAACTCAGAGGCCCAAGAAATGAGTGAAGTGCACATAAGTATA 1843
480LysAspValThrAspLeu.....LysValLysAla 489
1844 TGCAGAAAGATGTGCGACAGCTGGTGGAAATTAATTTCAATTTCTGTA 1893
490 ProValLeuAspGluTrpGlyLysGluPheThrAlaProValThrValLys 506
1894 CCATCTTCAAAAGATTTCACAG.....AA 1919
506 sValLeuAspLysAspGlyLysGluLeuLysGluLysLeuAlaL 523
1920 TATCATCGCGAAAGAGCGCAACATTAAAGATCCGTGAAGAGAGC. 1968
523 ysTrpValAsnArgGluLeuValLeuAsnAlaIaGlyGluAlaGly 539
1969AACACCAAGATTGACTTCTGTCA.....GAGAACAGCAAC 2004

540 AsnTrpThrValValLeuThrAlaLysSerGlyLysGluAlaLys.. 555
2005 TCTGAGACCATTAATTATCATCTGCGCAAGCAGCAACTGTGAGCTGCCG 2054
556AlaThrLeuAla...LeuG 561
2055 CAGCGGATTCGTCTATCCAGAAAGATCTGGCCACATAGCTGAAGTGG 2104
561 LuLeuLysAlaPro..... 565
2105 AGGTCTAATTCCTGCAAACTACACACTCTCTCATTTGGCACAAAGGC 2154
566GlyAlaPheSerLysPhe 571
2155 CGTGTGATCCGCTCCATCATGAGGAATGTGTGCTCCATATTCCT 2204
571 eGluVal.....ArgGlyLeuAspT 578
2205 TCTGTGGAAGTTCTGGAAGCGACACTGTGTCAATCAGGGGCTCTCT 2254
578 hrgLuLeuAspLys.....TyrValThrGluAsn 588
2255 CAGATGTGAAGAGCCACAGACAGCTCCTGCACTGGCTGAGAGAGAG 2304
589 GluLysAsnAlaMetThrValSerVal.....LeuProValAspAlaAs 603
2305 CAACCAAGAGTTTACTGTGACATCCGAGCCAGACAGAAATCCACAA 2354
603 nGlyLeuValLeuLysGlyAlaGluAlaIaGluLeuLysValThrThr 620
2355 ATTCCTATTGGCAAAAGGGGA..... 2376
620 hrasLysGluGlyLysGluValAlaAspAlaThrAspAlaGluValThrVal 636
2377 ..GGTAAATCCGAAGAGTGGCTGACAGTACTGAGGCCCGCATATATT 2424
637GluAsnAsnSerValIleThr...ValGlyGlu 647
2425 CCGGCTGCTGAAGATGAAGACCAAGACCTGATCAATCATAGAAAGGA 2474
647 yAlaLysAlaGlyLysThrThrLys..... 655
2475 GGATGTGTGCTGAGGCCCAAAAGAGCTGGAAGCATTTGATCCAGAAC 2524
656 .ValThrValValLeuAspGlyLysLeuIleThrThr.....HisSer 669
2525 TGGAGATGTGTGAGAGCATATGCTGTGGGACCCCAAAACACACCGC 2574
670 PheLysValValAspThrAlaProThrAlaLysGlyLeuAlaValGluP 686
2575 CATTTGTCAATCCGAAAGAGCAAGTTTACGAAATAATCTGTGAAGGTA 2624
686 eThrSerThrSerLeuLysGluValAlaIaProAsnAlaAspLeuLysAla 703
2625 TGGCGGGGTGTGTGTCAGC.....TTCCACGCTGTGACAGCGAGGTG 2668
703 IaLeuLeuAsnIleLeuSerValAspGlyValProIaThrThrAlaLys 719
2669 ATAAAGTCACTCTCAAGGTGCCAAGAGACTGTGAGGAGGCAACCAAAA 2718
720 AlaThrAlaSerAsnValGluPheValSerAlaAspThrAsnValValAl 736
2719 CGCATTCAGAAATCATTTGAGATCTGAAGCTCAATGACAGCGAATG 2768
736 a.....GluAsnGlyThrValGlyAlaLysGlyAla 747
2769 TGCTATATCCCAAGATTTCACCGATGTGATGAGGCCCAAAAGTTCA. 2817
747 hrSerIleTrpValLysAsnLeuThrValValLysAspGlyLysGluGln 763
2818AGATTCAG 2826

[illegible]

1848	AAAGCTCAAAAGAA...GCACTAGGCGTGCACGAAGAAAGCTTTAA	802
1849	sala...ValThrThrglnLysValGluValLysPheSerLysAlaValG	205
1850	: : : : : : : : : :	
801	GGCTAAATATATCTGCTAAAGAGCTCAAAAGAACTTTGCTAAAGGTTA	752
205	LulYsLeuThrLysGluAspLleLysValThrAsnLysAlaasnasp	221
751	AAAGCGTTAAGAAAATGAATCTAAATTTATCAAAAAG...AATAT	708
222	LysValLeuValLysGluValThrLeuSerGluAspLysArgSerAlaTh	238
707	AAAGAACTTGCTAGAAAGAGCGCCAAAGCTGCTAGAAAGAAAGCTTAA	658
238	rValGluLeuTyrSerAsnLeuAlaLysGlnThrTyrThrValaspv	255
657	GGCTAAATATATCTGCTAAAGAGCTGAAAG... : : : : :	627
255	aLaenLysValGlyLysThrGluValAlaValGlySerLeuGluAlaLys	271
626	...AACTTGCTAAAGAAATTTAAAGCAGTTAAAGAAAGATGCTTAA	582
272	ThrIleGluMetAlaAspGlnThrVal...AlaAspGluProThrAl	287
581	TTAATCAAAAAGACTATAGAAACTTACAAAGCTCTCAACAAGCTGC	532
287	aLeuGlnPheThrValLysAspGluasnGlyThrGluValValSerProg	304
531	TAGTAAAGTACCAAGAGAGTACTTAAACACT... : : : : :	498
304	LugLilGluPheValThrProAlaGluLysLleAsnAlaLysGly	320
497	...AAATTTGGAAAAAGAGCTGCAGTAAATTTAATAGAAAGCT	456
321	GluIleThrLeuAlaLysGlyThrSerThrThrValLysAlaValTyrLys	337
455	AAACTCAAAAAGAAAA...GTAATTCATTCTCA	421
337	sLysAspGlyLysValValaGluSerLysGluValLysValSerAlag	354
420	GAAG... : : : : : .GCTCA	410
354	LugLylAlaAlaValAlaSerLleSerAsnTyrThrValaLagLugLinsn	370
409	AGAAAGCTCACCAAGAGAGATTAAAGAACTGTAAG... : : : : :	363
371	LysAlaAspPheThrSerLysAspPheLysGluAsnAsnLysValTyrG1	387
362	... : : : : : .AAGTTATTTAAAGAAAGAAAGAAATTA... : : : : :	336
387	ugLysAspAsnAlaTyrValGlnValGluLeuLysAspGlnPheAsnAlay	404
335	...ATCAATTTACAAAAAAGATTC... : : : : :	314
404	alThrThrglyLysValGlyTyrGluSerLeuAsnThrGluValAlaVal	420
313	CAAAAGCTTCAAATTCAGAAAGAAAGAACTTTAAAGCAAAATTTCTCAA	264
421	ValAspLysAlaThrGlyLysValThrValLeuSerAlagLysAlaPr	437
263	GTCAAAAAAGGCTGCT... : : : : : .AAAGCTGTAAAGATC	232
437	oValLysValThrValLysAspSerLys	446
231	TGTTAGAAAGCCACTAAGAGTGTCAA	204
seq_name: gb_hlc:AK019969		
seq_documentation_block: 2413 bp mRNA RNC 05-JUL-2001		
LOCUS AK019969 8 days embryo cDNA, RIKEN full-length enriched		
DEFINITION Mus musculus 8 days embryo cDNA, full insert sequence.		
ACCESSION AK019969		

VERSION	AKO19969_1	GI:12860394
KEYWORDS	CAP trapper.	
SOURCE	Mus musculus (strain:C57BL/6J) 8 days embryo cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library clone:5730504C04.	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. Mus musculus	
REFERENCE	Garcinici, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)	
AUTHORS	1 (bases 1 to 2413) High efficiency full-length cDNA cloning Methods In enzymology. 303, 19-44 (1999)	
TITLE	2 (bases 1 to 2413) Garcinici, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome research. 10 (10), 1617-1630 (2000)	
JOURNAL	20499374	
MEDLINE	11042159	
PUBMED	3 (bases 1 to 2413) Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, Y., Nishik, K., Kitasuna, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, K., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multipipillary sequencer Genome research. 10 (11), 1757-1771 (2000)	
AUTHORS	4 (bases 1 to 2413) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 (bases 1 to 2413) Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Itawa, M., Kato, H., Kawai, J., Koizumi, Y., Komoto, H., Kouda, M., Koya, S., Kuhihara, C., Matsuyama, T., Miyazaki, A., Nishik, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sobabe, Y., Suzuki, H., Tagami, M., Tadawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyai, T., Yamawara, Y., Yasunishi, A., Yoshida, K., Yoshihino, M., Muramatsu, M., and Hayashizaki, Y. Direct Submission Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome_research@riken.go.jp, Fax: 81-45-503-9216) URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.	
COMMENT	Location/Qualifiers 1..2413 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /db_xref="MGD:MGI:1911416" /db_xref="MGD:MGI:1924829" /clone="5730504C04" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="8 days embryo" 783 a 500 c 798 g 332 t	
FEATURES		
SOURCE		
PAGE COUNT	783 a 500 c 798 g 332 t	

ORIGIN	Alignment_scores:	Quality:	135.00	Length:	515
	Ratio:	0.511	Gaps:	24	
	Percent Similarity:	51.262	Percent Identity:	21.942	
Alignment_block:					
05-09-754-947-1 x AK019969	..				
Align seg 1/1 to: AK019969 from: 1 to: 2413					
151	GLYTHRGLYASPGINTPRGLUPROLYSSTRVALTHRLYSALAGLUAL	167			
173	GGAGGAGGCGAAGCGCAACCGATCTCACAAATGAGACAAGACAGA	222			
167	AALA GLAPHELLEALALSTRHSPYSGINPHEGLYTHRGUALLAALA	183			
223	TGCAGCGCCACATTGAGGACCTAGAAAGACACATCGATGAGAGC	266			
184	LYSVALGLUSERALALYSALVALTHRTHGLINYSVALGLUVALYSPH	200			
267GAGGGGGCCCG.....CAAAAGCTGCAGCTGGAGAA	298			
200	ESERLYSALVALGLULYSLEUTHRLYSGLUASPDILLYSVALTHRASN	217			
239	GGTGACAGCAGAGCTTAATATCAAGAGATGGAAGAGAGGTCTGCTTC	348			
217	YSALASNAASAPLYSVALLEUVALYSGLUVALTHRLEUSERGLUASP	233			
349	TGCAAGACACAGATTCCAAATTATCAAGAAAGAAACTCATGAGAGAC	398			
234	LYSARGSERALATHVALGLULEUTYSSERISNEUALALALYSGLINTH	250			
399	CGA.....ATTGCTGAGTGTCTCTCAGCTGCTGAAGAGAGAA	439			
250	ITYRTHRVALASPVALLALYSVAL..GLYLYSTHGLUVALALVALG	266			
440	AAGGCAAAAAAAGCTGGCCCAAAATCAGSAAATACCAAGAGTGAATCT	489			
266	LYSERLEUGLUALALYS.....THRTLE 273				
490	CGGACTTGAGAAAGCGCTTGAAAGAGAGAGAAACCTGCACAGCACTG	539			
274	GLMETALA.....ASPDLPNTHVALVALALASPGI 284				
540	GAAAAGGCCAAACGGAAGCTGATGGGAAACAACCATCTGCAGACCA	589			
284	UPROTHRALALEUGINPHEHTRVALLYSASPGULASNGLYTHRGUVALY	301			
590	GATGCTGAGCTGCAGGACGACAGTC..GATGAGCTCAAAAGTCCAGTTGA	636			
301	ALSERPROGLUGLYTLEGLNPEVALTHRPROALAAGLUYSILEASN	317			
637	CCAAAGAAAGAG.....GAGGAGCTTCAG 659				
318	ALALYSGLYGLUIETHRLEUALALALYSGLYTHSERHTRHVALYSAI	334			
660	GGG.....GCGCTGGCCAGAGATCATAGACA..... 689				
334	AVALTLYRSLYSASPGIYLYSVALVALALAGLUSERLYGLUVALLYSV	351			
690	CTGACACAGAATATATGCACCTTAAGTTGCACGG.....GAGC 726				
351	ALSERALAGLUYALALALVALALASERTILESERASNTPRTHVALAIA	367			
727	TGCAGGCCCAATGCGAGAGCTCCAGAAAGACTTGGTCTGAAAAGCGT	776			
368	GLULINASNLYSALASAPPHETHSERLYSASPHELYSLINASNL	384			
777	TCAAGGAACAGCGTGAAGAAACAAAACGGGAGCTTGAGTGAAGAGCTGG	826			

```

384 svaLlytGluGlyspAsnAlaTyrValGlnValGluLeuysAspGlnP 401
      ::::: ::::: ||||| ::::: |||
827 AGCT.....CTGAAGACAGAGCTGGAGCCACC 855
401 heaSnAlaValThrThrGlyLysValGluTyrGluSerLeuAsnThrGlu 417
      ::::: ::::: ::::: ::::: ::::: |||
856 TAGACACCACGACGCTCAGCAGGAACCTCCGACAAACGTGACGAGAA 905
418 ValAlaValAlaAspLysAlaThrGlyLysValThrValLeuSerAlaG1 434
      ||||| ::::: ||||| |||||
906 GTGGCAGAGCTGAAGAAGCT.....CTTGAGCATGA 937
434 yLysAlaProValLysValThrValLysAspSerLysGlyLysAlaLeuV 451
      ::::: ::::: ||||| ::::: :::::
938 AACTAAGAACCCAGAACCTCAGATCCAGGACATGAACAGAG..... 980
451 alSerH1sthrValGluLeuAlaPheAlaGluLysAlaMetLys... 466
      ||||| ::::: ||||| ::::: :::::
981 .....CATGCCACAGCGCTGGAGGAGCTTTCCGAGCAGCTGGAGCAACG 1025
467 .....AspLeuLysLeuGluLysThrAsnValAlaLeuSerThrLy 480
      ::||| ::||| ::||| ::||| ::|||
1026 AAAAGGTTCAAAAGCCAACTGAGAGAAGAACAAACAGGCGCTGAGACA.. 1073
480 sAspValThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyL 497
      ||| ::||| ::::: ||||| ::|||
1074 .GACAAACAGAGAGCTGCGTGTAGGTGAAGGTGCGTGCAGCGAGTAAGG 1122
497 ysgLupHethrAlaProValThrValLysValLeuAspLysAspLys 513
      |||
1123 CGGAG.....TCA 1130
514 GluLeuLysGluGlnLysLeuGluAlaLysTyrValAsnArgGluLeuVa 530
      ||| ||||| ::||| ::||| ::||| ::|||
1131 GAGCAACAAGAGAGAGAGAGCTGATGCCACAGCTCCAGAG..... 1169
530 lLeuAsnAlaAlaGlyGlnLysAlaGlyAsnTyrThrValLysLeuThra 547
      ||||| ::||| ::||| ::||| ::|||
1170 .CTCCATCCCAAGGTGCAGAGGGTGCAGGCTCAGGCTCAGGAGTGGCCG 1218
547 lAlysSer.....GlyGluLysGluAlaLysAlaThrLeuAla 559
      ||||| ::||| ::||| ::|||
1219 AGAAAGCAACACACTACAGAAATGACCTGATATGTCACACCTGCTG 1268
560 LeuGluLeuLysAlaProGlyAlaPheSerLysPheGluAlaArgGlyLe 576
      ||| ::||| ||||| ::|||
1269 GAAGAGAGCTGAGAGAAAGATTAACTTTGGCAGAGATGACGCTGGCT 1318
576 uAspThrGluLeu.....AspLysTyrValThrGluLysGln 590
      ||| ::||| ||||| ::|||
1319 CGAGTCTCAACTACAGGACACACAGGAGCTCCTTCAGAAAGACACAGCG 1368
590 yAsn.....AlaMetThrValSerValLeuProValAspAlaAsn 603
      ::||| ::||| ::|||
1369 AGAAACGTAACTGACGATCGGATCCGCGCAGGAGGAGAGAAAGAAC 1418
604 GlyLeuValLeuLysGlyAlaGluAlaAlaGluLeuLysValThrThr 620
      ::||| ::||| ||||| ::|||
1419 AGCCTTCAGGAGCAGCAGGAGGAGGAGAGAGCCAG.....AA 1459
620 rAsnLysGluLysGluValAspAlaThrAspAlaGlnVal 634
      ||| ||| ||||| ||| ::||| ::|||
1460 GAACCTGAG...AAGCAGGTGTGGCTCTGCAAGTCCACAGTGG 1499

seq_name: gb_hlc:AK019450

seq_documentation_block: 2633 bp mRNA HMC 05-JUL-2001
LOCUS AK019450
DEFINITION Mus musculus 14, 17 days embryo head cDNA, RIKEN full-length
enriched library, clone:3230402H02, full insert sequence.
ACCESSION AK019450
VERSION AK019450.1 GI:12859660
KEYWORDS CAP trapper.

```

bulk excision from lambda FLC I. Cloning sites, 5' end: SalI; 3'
end: BamHI. Host: DH10B.

FEATURES

Location/Qualifiers
source

1..2633
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGI:1911407"
/db_xref="MGI:1924822"
/clone="3230402H02"
/issue_type="head"
/clone_lib="Riken full-length enriched mouse cDNA library"
/dev_stage="14, 17 days embryo"
172..2343
/note="putative"
/protein_id="BAB31726.1"
/db_xref="GI:12859661"
/translation="MGFCNSSRNQOMLSTEDGKFLIKSGCLGIGISNSSRPPQPAV
ATPCADAPRWCHAOEGLFLEVENTSLFKQKHKKVVKISKYLDWMKLVINKBGR
ANESLCLQAGLGEVSVTARNSASQIATANTVPSPSHSNTTEFLQSTETL
RSTAEIIGSTAEIIRNTAKTIGTAEIIRNTAKTIGTAEIIRNTAKTIGTAEIIR
TAKTIGTSEAYSSOSSKRGRLPHLHAGATDESMSPILTPPSSITTEGVABOVKN
FTLESRVSSLSASIQMRTPASPCNPSLIYSDTSGPMCHPILRIDNFTYGGNKDLO
ACTVYRIRIVSLDGEESTVLQTDPLPARFENSKTSTLQVWMTSSSKVSYE
VOLFDNDKIDQVOVESTVTSOTTFLENGSTKVAITVSGEKRSFPVYINST
VPSPVKDLGISPPNSLLISWSRSGNVEOYRLVLDKDAIVQDNVDRDVSVAHE
LPSGHLNLTIVTMSGLONSRLKLVRTAPMEVSNLVNDRLSLNKKMCKPQDQ
DEYSITLSHOGTIEKSTLAPVTEFOFKDLVGRLOYTICISGELSAEKSAAGRT
VPEKVRINLVNEMKSFYVMTWTPAGDMHYRIYLFNESIYLVNTITVKGKETHAL
DGLFLIRGQYEIVVESGNLRNSRCGRGRKSNLIORT"

CDS

BASE COUNT 718 a 665 c 643 g 607 t
ORIGIN

alignment_scores:

Quality: 134.50 Length: 706
Ratio: 0.388 Gaps: 32
Percent Similarity: 49.150 Percent Identity: 18.839

alignment_block:

US-09-754-947-1 x AK019450 ..

Align seg 1/1 to: AK019450 from: 1 to: 2633

146 LeuGluIleSerValGlyThrGlyAspGlnTrpGluProLysLysThrVal 162
|||||
1259 CTGGGTATCTCCAACTCTCAAGGGCCCTTCCACACGACGATTGGCTAC 308
162 LThrLysAlaGluAlaGlnPheIleAlaLysThrAspLysGlnPheG 179
: |||||
309 TCCTGTGCCCAGGACCCCGATGACCTGCCATGCTCAGGAAGGTTTC. 357
179 LThrGluAlaAlaLysValGluSerAla.....LysAlaVal 191
358CTTGAGGTGGAAAAATACCTCAGCTTTCTTCAAGAAACAA 396
192 ThrThrGlnLysValGluValLysPheSerLysAlaValGluLysLeuTh 208
::: |||
397 AACCATTAAGTAGTGTCAAGACATCATGTAATACCTCGATTACTCGAT 446
208 rLysGluAspLleLysValThrAsnLysAlaAsnAsnAspLysValLeuV 225
||| |||||
447 GAAGTATGATTATCAACAAGAGGAGACTGCCAATGAAGACCTCGTGC 496
225 aLlysGluValaLThrLeuSerGlnAspLysArgSerAlaThrValGluLeu 241
: |||||
497 TGAAGCAAGCTGGCTGGTACAGAA.....GTCTCGCTCGGAGCT 537
242 TyrSerAsnLeuAlaAlaLysGlnThrTyrThrValAspValAsnLysVa 258
||| |||||
538 GCTGAAAACTAGCTCTCCACAGATTGCCACTACT...GCTAACACCGT 584

258 L.....GlyLysThrGluValAlaValGluLys 267
+ |||||
585 CCCATATAGCCCAAGCCACATTACCATATACAGAGACATTCCTCGGGA 634
267 erLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAsp 283
||| |||||
635 GC...ACTACAGAGACCTCAGAGACCTCAGAGACCTCGGGAGACT 681
284 GluProThrAlaLeuGlnPheThrValLysAspLysnGlyThrGluVal 300
||| |||||
682 GCAGAGACC...CTCAGAGACACTGCAGAGACCTCGGAGACT.... 723
300 lValSerProGluGlyIleGluPheValThrProAlaAlaGluLysIleA 317
724GCAGAGACCTCAGAGACCTGCAGAGACCTCGGAGACT 757
317 snAlaLysGlyGluLe.....ThrLeuAlaLysGlyThrSerThr 331
758 GGACACTGCAGAGACCTCAGAGACCTGCAGAGACCTCGGAGACT 807
332 ValLysAlaValTyrLysLysAspGlyLysValAlaLysLysGly 348
808 GCAAGACCCCTC..... 819
348 uValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTrp 365
820GGAGACCTGCAGAGACCTCGGAGGAGC.... 846
365 hrValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
847ACTTACAGAGCCTACAGTCAG 867
382 AsnAsnLysValTyrGluGlyAspAsnAlaTyrValGluValGluLeu 398
868 AGCAGAGACCAAGAGAGCTCCCTCCATCGACAGAGCTGAGCTACAGA 917
398 sAspGlnPheAsnAlaValThrThrGlyLysValGluTyrGluSerLeu 415
918 CGAATCTGTGGAGCCGCTGACT....CCACCTTCTCCGAGCTCA 961
415 snThrGluValAlaValAlaLysPheLysValThrGlyLysValThrValLeu 431
962 CCACAGAGACTGGAGTGGCAGAGCACTGAATGTAATCTTACACTGCTTG 1011
432 SerAlaGlyLysAlaProValLysValThrValLys..... 443
1012 GAGTCCAGAGCTCTGACTGTGTGACGCTTATCCAGAGGAGACTTTCG 1061
444AspSerLysGlyLysA 449
1062 GTCAACCTGTACTTATGCTCATCTACAGCAGTGAATCCTCGGG...C 1108
449 lAlaLeuValSerHisThrValGluIleGluAlaPheAlaGlnLysAla... 464
1109 CCAATGTCGTCATCTTATTCGATATACACAATTTACTTACGATGTATAC 1158
465 MetLysAspLleLysLeu..... 470
1159 CCCAAGATTTTACAGAGAGGAGCCGCTATATACCTCAGATTGTTTCTCT 1208
471GluLysThrAsnValAlaLeuSerThrLysAspValThrAspL 485
1209 GAGTGAAGAGAGAGCACTGTGCTTACAGACAGATCCGTTG..... 1251
485 euLysValLysAlaProValLeuAspGlnTyrGlyLysGluPheThrAla 501
1252CTTCGTCAGAGTTTGAAGTCAATTCGGAGAAACAGCA 1290
502 ProValThr..... 504
1291 TCACACACCTGCAGAGTCGCGTGCATCCTCTTCTGGAAGATCTCTCG 1340
505ValLysValLeuAspLysAspGlyLysGluLeuLysGlu...G 518

```

1341  GTATGAGGTGCAACATATTGATGATCATTAACGATCAAAAGATACAAAGAACTCC 1390
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
518  InLysLeuGluAlaLysTYRValAsnArgGluLeuValLeuAsnAlaIle 534
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1391  AAGTTCAAGAAAGATGCACACCTGGAGCCAAATATACTTTCTGAACTTCACCT 1440
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
535  GLyGInGluAlaGlyAsnTrpTrpIleValLeuThrAlaLysSerGly 551
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1441  ....GAGGGTAACGTTTACAAAGTTGGCATCAGCTCTTTCTGGGAGA 1484
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
551  uLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyLap 568
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1485  AAAGGCC.....TCTCTTCCGGGTATATCAATAGGCTCTA 1519
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
568  heSerLysPheGluValArgGlyLeuAspThrGluLeuAspLysTYRVal 584
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1520  CAGTACCATCTCCAGTGAAGAAGATCTTGGCAT..... 1551
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
585  ThrGluLysAsnGluLysAsnIleMetThrValSerValLeu..... 598
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1552  ....TCCCCCAATCTTAATCTCTCTTAATTTCTCGGTCTGCTGCTTC 1595
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
599  ...ProValAspAlaAsnGlyLeuValLeu.....LysGlyAlaGlu 612
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1596  TGGAGATGGGAACATACAGCTGGTGCTAAAGATAAAGAGGGCC... 1641
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
612  IaIaGluLeuLysValThrThrAsnLysGluLysGluValAsp 628
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1642  ....ATCGTTCAAGACACAAAGCTGGACAGGCGGTGAATCTCT 1680
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
629  AlaThrAspAlaGluValThr.....ValGlnAsnSerValIle 642
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1681  TATGCTTTTACAGAGCTGAGCCCTGGCCACCTCTACAACTCAGTATTGT 1730
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
642  eThrValGlyGInGlyAlaLysAlaGly..... 651
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1731  CACCATGGCGCTGGGAGCTGCAGAAACCTGCAGAGTGGAACTGGTGAAGACGG 1780
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
652  ....GluThrTYRlysValThrValValLeuAspGlyLysLeuIle 665
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1781  CTCCCATGGGAAGTCTCAAACTGAAGGTGCAAAATGACAGGAGGTGACC 1830
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
666  ThrThrHis.....SerPheIy 671
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1831  TCTCTAAATGTGAAGTGCAGAAACCCCTGGGAGATGATATTCTCTACG 1880
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
671  sValVal.....AspThrAlaProThrAlaLysGlyLeuAlaVal 684
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1881  CATTAACCCGTCTCACCAAGGAGACATCAAGAATCCAAAAACATTATACAC 1930
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
684  aGluPheThrSerThrSerLeuLysGluValAlaProAsnAlaAspLeu 700
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1931  CTCTGTTTACTGAACATCAATTAAGAGACTAGTCCCTGGA..... 1971
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
701  LysAlaIaIaLeuLeuAsnIleLeuSerValAspGlyValProIaIaThr 717
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
1972  CGGCTTTACCAAGTGACCATCAGCTGCATCTTGgt.....GAGCTCTC 2015
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
717  IaIaLysAlaThrIaIaSerAsnValGluPheValSerIaIaAspThrAsnV 734
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
2016  TGCTGAGAGTCAAGCAGCGGGGAGAACAGTTCCAGAAAAAGTGAGGATTC 2065
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
734  aIValAlaGluAsnGlyThrValGlyAlaLysGlyAlaIaThrSerIleTYR 750
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
2066  TGGTTTCCACAAAC.....GAGATTGG 2088
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
751  ValLysAsnLeuThrVal 756
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
2089  ATGAAGTCTCTTACAGATG 2106

```

seq. documentation block:	2582 bp	mrna	HTC	05-jul-2001
LOCUS	AK016539			
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:493243JP20, full insert sequence.			
ACCESSION	AK016539			
VERSION	AK016539.1	GI:12855324		
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA library clone_493243JP20.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 2582)			
TITLE	Carninci, P. and Hayashizaki, Y.			
JOURNAL	High-efficiency full-length cDNA cloning			
MEDLINE	Methods in enzymology. 303, 19-44 (1999)			
PUBMED	99279253			
REFERENCE	10349636			
AUTHORS	2 (bases 1 to 2582)			
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
MEDLINE	Genome research. 10 (10), 1617-1630 (2000)			
PUBMED	20499374			
REFERENCE	11042159			
AUTHORS	3 (bases 1 to 2582)			
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagata, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, T., Nishi, K., Kikunari, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.			
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format			
MEDLINE	sequencing pipeline with 384 multicapillary sequencer			
PUBMED	Genome research. 10 (11), 1757-1771 (2000)			
REFERENCE	11076861			
AUTHORS	4 (bases 1 to 2582)			
TITLE	The RIKEN Genome Exploration Research Group Phase II Team and the			
JOURNAL	FANTOM Consortium.			
MEDLINE	Functional annotation of a full-length mouse cDNA collection			
PUBMED	Nature 409, 685-690 (2001)			
REFERENCE	5 (bases 1 to 2582)			
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanaagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirabe, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shikata, T., Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toy, T., Yamamuro, T., Yasunishi, Y., Yoshida, K., Tejima, Y., Toy, T., Yamamuro, T., Yasunishi, Y., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of			
REFERENCE	Physiological and Chemical Research (RIKEN), Laboratory for Genome			
AUTHORS	Exploration Research Group, RIKEN Genomic Sciences Center (GSC),			
KEYWORDS	RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,			
SOURCE	Kanagawa 230-0045, Japan (E-mail:genome_research.riken.go.jp,			
URL	URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,			
COMMENT	Fax:81-45-503-9216)			
	Please visit our web site (http://genome.gsc.riken.go.jp/) for			

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5'-GAGAGGAGCAAGCATCCAGACGCTCTTTTTTTTTTTTTTAA 3'], cDNA was

KLuyveromyces marxianus, genomic survey sequence.
AL423584
VERSION AL423584.1 GI:12206778
KEYWORDS GSS.
SOURCE KLuyveromyces marxianus.
ORGANISM KLuyveromyces marxianus.
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
AUTHORS 1 (bases 1 to 943)
Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.
and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 12.
JOURNAL Kluyveromyces marxianus var. marxianus
MEDLINE FEBS Lett. 487 (1), 71-75 (2000)
REFERENCE 20584722
AUTHORS 2 (bases 1 to 943)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brotter,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Olier-Kalogeropoulos,O., Potier,S.,
Saurin,M., Tekala,F., Toffano-Mioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
AUTHORS 3 (bases 1 to 943)
REFERENCE Direct Submission
TITLE Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 Evry cedex, FRANCE. (E-mail :
sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
1..943
 /organism="Kluyveromyces marxianus"
 /strain="CBS 712"
 /variety="marxianus"
 /db_xref="taxon:4911"
 /clone="AZ0AA009C10"
 /clone_11b="AZ0AA"
 /note="end : T3"
 <1..>933
 /note="similar to Saccharomyces cerevisiae ORF YJL176c [
 SW13 : transcription regulatory protein]
 1 putative frameshift(s)"
 /evidence="not_experimental"
BASE COUNT 388 a 188 c 208 g 159 t
ORIGIN

alignment_scores:
 Quality: 131.50 Length: 414
 Ratio: 0.678 Gaps: 18
Percent Similarity: 46.860 Percent Identity: 22.464

alignment_block:
US-09-754-947-1 x CNS070GA ..
Align seg 1/1 to: CNS070GA from: 1 to: 943

128 LysAspLeuGluThrLeuAsnTrpGly.....LysG1 138
|||||||
1 AAGACCTTGCGGTAAATTTGGTTCCCATATGCATTCATTAATATC 50

138 uLysAlaAsnIleLeuValGluGlyIleSerValGlyThrGlyAsp. 154
: |||||
51 TGATATATCCAGTTATGAGCACTTGGCATCTTGTATGCTGTGTTATATC 100
155GlnTrpGluProLysThrValThrLysAlaGluAla 167
|||
101 CTAAATATTTGCCAAGAAATTGCATCAAGAGCGCAATGCATTCAGACGAG 150
168 AlagInPheIleAlaLysThrAspLysGln.....PheGlyThrG1 181
|||
151 AAGCAA.....AGACTGATACGAGGTAAACCATATTCATACGA 191
181 uAlaAlaLysValGluSerAlaLysAlaValThrThrGlnLysValGluV 198
|||
192 AGAAGAAAGAAATACGAAACCTAACAACGTGCGAGAAAGACGACGAG 241
198 alLysPheSerLysAlaValGluLysLeuThrLysGlu.....AspIle 212
|||
242 ACAAGGAGAAAGACGAGAACAGAACAGACGAGGAGAAACAGACGAG 291
213 LysValThrAsnLysAlaAsnAspLysValLeuValLysGluValThr 229
|||
292 AAGGAGACAGAAAGCGGAGAGAGAAAGAGACAGAAAGCATATAGA 341
229 rLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsnLeuA 246
::: ||||| ::: ||| |||
342 AATG...GTTGATTAACCTGAAGAAACGAAAGAG..... 372
246 lAlaLysGlnThrTyrThrValAspValAsnLysValGlyLysThrGlu 262
:::
373 ..GCCGAAATGTAAACGATTCCAGACGTCGAGAAAGCGATCTGAA 420
263 ValAlaValGlySerLeuGluAlaLys...ThrIleGluMetAlaAspG1 278
|||
421 GAACAAGAAAGATTCCATGAGCGCAAAAGAGAGCTGAAATGCTGTATCA 470
278 nThrValValAlaAspGluProThrAlaLeuGlnPheThrValLysAsp 295
|
471 A.....GACCAAAAGCC...CAGACTCTACATCAAGATA 502
295 LysnGlyThrGluValValSerProGluLysIleGluPheValThrPro 311
::: ||||| |||||
503 GAAT.....AGCCAGAGAGATAGAA..... 525
312 AlAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLysGlyTh 328
|||
526AAATCCAC 533
328 rSerThrThrValLysAlaValTyrLysLysAspGlyLysValAlaLag 345
|||
534 ATCTCCAAAGGCCAATGACGTCGAAAGAAAGAT..... 567
345 lSerLysGluValLysValSerAlaGluLysAlaValAlaLaserIle 361
567 567
362 SerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLysAs 378
567 567
378 pPheLysGlnAsnAsnLysValTyrGluGlyAspAsnAlaTyrValGlnV 395
|||||
568AATTAATGAAGCCAAAGATGAAGCCGAAAGCTGATTAAGAA 607
395 alGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluTyr 411
::: |||
608 CTCAGCTTAAGACTGAACCGAAACTGAACCTCAACCGAACCCAAACC 657
412 GluSerLeuAsnThrGluValAlaValAspLysAlaThrGlyLysVal 428
|||
658 GAA.....ACCGAATCAAGCCCGTATCTGAAGCT..... 687

428 lthvalleuseralaglylsalaprovalthrsvalthrsasps 445
688
445 ertylsglylsalaleuvalserthrsvalguilegualaphala 461
707 CCACCGAAGTACCTTGGCCACTTGTGCTCCGCTCACACGATTTCGCT 756
462 GluLysAlaMetLysAspIleLysLeuGluLysThrAsnValAlaLeuSe 478
757 ACAAAC.....CAGGAGACCTCATGACACAGCACAATGACCTAAT 800
478 rthrsaspvalthrspleuLysValLysAlaProvalLeuaspclnt 495
801 TACCACCACTCACTCAAGGCTGATGTAATGAAAATCTACACCAA 850
495 yrglylsygluphethralaprovalthrsvalleuaspLysasp 511
851 TGCAGAAA.....TCACGTGAATTAGACGGAAGCTATTTCATAGCAAG 894
512 gLyLysGluLeuLysGluLysLeuGlu...AlaLysTyr 524
895 CAGAGACATATCTTCATTCAAGACTAAGTCTTGCAAAATAT 936
seq_name: gb_hc:BC008697

seq_documentation_block:
LOCUS BC008697 2056 bp mRNA HTC 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3342723, mRNA.
ACCESSION BC008697
VERSION BC008697.1 GI:14713897
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2056)
Strasberg,R.
Direct Submission
Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Pirahvu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Nastasia van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRML Plate: 22 Row: J Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 1944184
This clone has the following problem: incomplete processing.

FEATURES
Source
1..2056
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3342723"
/tissue_type="Eye, retinoblastoma"

/clone_lib="NIH_MGC_16"
/lab_host="DH10B-R"
/note="Vector: pORF7"
BASE COUNT 773 a 332 c 479 g 472 t
ORIGIN
alignment_scores:
Quality: 131.00 Length: 695
Ratio: 0.424 Gaps: 33
Percent Similarity: 44.460 Percent Identity: 18.849
alignment_block:
US-09-754-947-1 x BC008697 ..
Align seg 1/1 to: BC008697 from: 1 to: 2056

181 GluAlaAlaLysValGluSerAlaLysAlaValThrThrGlnLysValG 197
347 GATGCTGTGATGCTTGAACTTGAGAGACCATTAAGTGAAGTCCCAATACA 396
197 uValLysPheSerLysAlaValGluLysLeuThrLysGluAspIleLysV 214
397 AGTA.....GTTGCAAGCAGACTTTAGATG 422
214 aThrAsnLysAlaAsnAsnAspLysValLeuValLysGluVal..... 228
423 CATTAACCTACTTACATGATTAATTAAGATCATCCACAGAGATCGAAGGCT 472
229ThrLeuSerGluAspLysArgSerLathrValG 240
473 GGCAACATTCCTTACCTTATGATGAGATATCAAAATGGCGGATTTT.. 520
240 uLeuTyrSerAsnLeuAlaAlaLysGlnThrTyrThrValAspValAsn 257
521GGAGTATCAGCTTAAAAACAGAGACATTCMA..... 553
257 ysValGlyTyrGluValAlaValGlySerLeuGluAlaLysThrIle 273
554AAGAGAGATTCCTTATGTTGTTACA.....CCATAT 583
274 GluMetValAspGlnThrValValAlaAspGluProThrAlaLeuGlnP 290
584 TGGATGGCTCCTGAAGTAGTATGTGAAACATCTAAGACAGACCCCTA 633
290 eThrValLysAspGlu.....AsnGlyThrGluValValSerPro 304
634 TGACTACAAAGCTGATGTTGGTCCCGGTATCATCTTAATGAATAG 683
304 LucGlyIleGluPheValThrProAlaAlaGluLysIleAsn..... 317
684 CTGAGATGAA.....CCACCTCATCATGATTAATTAATCCAAATGCGA 724
318AlaLysGlyGlu...IleThrLeuAlaLysGlyTh 328
725 GTGCTGCTAAAAATAGCAAAATCTGAGCCACCTTACATTACACAGCCATC 774
328 r.....SerThrThrValLysAlaValTyrLysLysAspGlyLysVal 343
775 CAGATGCTCTCAAAATTTTAAGGACTTCTTAAGAAA..... 811
343 aAlaGluSerLysGluValLysValSerAlaGluGluAlaAlaValAla 359
812TCCTTAAGAAAGAAATGTGATGCC..... 835
360 SerLeuSerAsnProThrValAlaGluGluAlaLysLysAlaAspPheThr 376
836AGGTGACTACATCTCAAGCTCTGACGATCATCCCTTTGATC 876
376 rLysAspPheLysGluAsnAsnLysValTyrGluGluLysAspAsnAlaTyrV 393
877 TGTTCAT.....TCCAAACAACCATCCGAGAAATTCATTGCGAGAG 917

```

393 aGlnValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysVal 409
      :::::||||: ::::: ||| |||||
918 CCAAGCTGGAAGTAACAGAGAA.....GTGGAAGATGCCAAGAG 958
      ||| |||||: :::::|||||
410 GluTyrGlnSerLeuAsnThrGluValAlaValAlaLysAlaThrG1 426
      ||| |||||: :::::|||||
959 GAAGATGAGAGAGAGAAACAGAA.....AA 984
      :::::||||: :::::||||:
426 yLysValThrValLeuSerAlaGlyLysAlaProValLysValThrVal 443
      ::::: ||: ::::: |||||
985 TTCTCTGCCAATACCTGCAGAGTAAGCGTGCAATCTTGACCTAGTATCG 1034
      ::::: ||: ::::: |||||
443 yAspSerLysGlyLysAlaLeuValSerHisThrValGluIleGluAla 459
      ::::: ||: ::::: |||||
1035 CCAGCTCGAAGAGATAACTTTCACAAATGCTGTATTGAGAGTCT 1084
      ::::: ||: ::::: |||||
460 PheAlaGlnLysAlaMetLysAspIleLysLeuGluLysThrAsn...Va 475
      ::::: |||||: ::::: |||||
1085 CTCCTGAGAAACAGAACGTAGTACTCTGACAGTTAACTCACACAGCAA 1134
      ::::: |||||: ::::: |||||
475 lAlaLeuSerThrLysAspValThrAspLeuLysValLysAla..... 489
      |||||: ||: ::::: |||||
1135 AATCTTAATCAAAAACCCACCTGATGAACCTGAAAAGCTGTGAGAG 1184
      ::::: |||||: ::::: |||||
489 ..... 489
1185 ATATTAAATGAACATATTACCGATGCTCAGTTAGAACAATGACTGAACCTC 1234
      ::::: |||||: ::::: |||||
490 .....Prova 491
1235 CATGACAGAACAGCATATCAAGAGAAATGAAGAGAGAGAGGCCAA 1284
      ::::: |||||: ::::: |||||
491 lLeuAspGlnTyr..... 495
      |||||: |||||:
1285 GCTTGAAATCTGCCTGCACAGAAAGACCAAACTGTGACATTATT 1334
      ::::: |||||: ::::: |||||
496 .....GlyLysGluPheThrAlaProValThrValLysVal... 507
      |||||: |||||: |||||:
1335 CAGTCAGTGAAGAAAGAGAAATATATATATATTAACCTTAGAAAACAAAT 1384
      ::::: |||||: ::::: |||||
508 .....LeuAspLysAspGlyLysGluLeuLysGluG1 518
      ::::: |||||: ::::: |||||
1385 ATTGACATATCTAAATCTGAGAGAAAGACATCAGAAAGCAACA 1434
      ::::: |||||: ::::: |||||
518 nLysLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAlaAlaG 535
      ::::: |||||: |||||: |||||:
1435 GATGTTTGAAATTAAGCTTATTAATCTGAGAAATTAAGATACTATT 1484
      ::::: |||||: ::::: |||||
535 yGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLysSerGlyGlu 551
      |||||: |||||: |||||: |||||
1485 TGCAAACGATAGAT.....TTAGTTCTCAAGAGACTGTGAGAA 1522
      ::::: |||||: ::::: |||||
552 LysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPh 568
      |||||: |||||: |||||:
1523 AAACAGCA..... 1531
      ::::: |||||: ::::: |||||
568 eSerLysPheGluValArgGlyLeuAspThrGlu.....LeuAspLysT 583
      ::::: |||||: |||||: |||||
1532 .....AATATTCAAGCAGTTGATAGTAGGAGTTGGCTTACAAGG 1571
      ::::: |||||: ::::: |||||
583 yrValThrGluGlu.....AsnGlnLysAsnAla 592
      |||||: |||||: |||||:
1572 AAGACACCCCAAGAAATTGGGAGAGACGACAAACCAAAAGATGTG 1621
      ::::: |||||: ::::: |||||
593 MetThrValSerValLeuProValAspAlaAsnGlyLeuValLeuLysG1 609
      ::::: |||||: ::::: |||||
1622 .....ATCAGCAATACAAAGTGTGATAGGAAAC 1650
      ::::: |||||: ::::: |||||
609 yAlaGluAlaAlaGluLeuLysValThrThrAsnLysGluLysG 626
      ::::: |||||: |||||: |||||
1651 ATGTGAGGACACAGATGTG.....GCTCAGA 1676
      ::::: |||||: ::::: |||||
626 lValAlaPheAlaThrAspAlaGlnValThrValGlnAsnAsn.....Ser 640
      ::::: |||||: ::::: |||||

```

```

1677 AAGTGAATGAAGACAGTGTGATGATACGACAGATTAATGATGGGAAAGAA 1726
      ::|||::: ::|||::: ||| ::|||::: ::
641 ValIleThrValGlyGln.....GlyAlaLys 649
      |||||: |||||: |||||:
1727 GTGCTCCAGTAGGCCAGAAATTAATTATTAAGCCCATGGGGTCTCTGA 1776
      ::|||::: ||| ::|||:::
649 sAlaGlyGluThrTyrLysValThrValValLeuAspGlyLysLeuIleT 666
      ::|||::: ||| ::|||:::
1777 GCGTGTGTGCTTAAGGAAGTTCTATTAA..... 1807
      ::|||::: ||| ::|||:::
666 hTrpHisSerPheLysValValAspThrAlaProThrAlaLysGlyLeu 682
      ::|||::: ||| ::|||:::
1808 .....GAATAGTGAATGAATGAATGAATGAAGAGGT... 1840
      ::|||::: ||| ::|||:::
683 AlaValGluPheThrSerThrSerLeuLysGluValAlaProAsnAlaAs 699
      ::|||::: ||| ::|||:::
1841 .....AAAAATAGGAAACACAAATAAACACTTC 1869
      ::|||::: ||| ::|||:::
699 pLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValProAlaT 716
      ::|||::: ||| ::|||:::
1870 A.....GAGAACATATGACATCATGAGAGAACAGCAA 1904
      ::|||::: ||| ::|||:::
716 hTrp.....AlaLysAlaThrAlaSerAsn 724
      ::|||::: ||| ::|||:::
1905 CAACCTGAAGGTGAAGAAATCACTAGTCAGTCAAGTACACTGAAGAAATGAG 1954
      ::|||::: ||| ::|||:::
725 ValGluPheValSerAlaAspThrAsnValValAlaGluAsnGlyThrVa 741
      ::|||::: ||| ::|||:::
1955 GTCGAAGGTGTGTGCTGATCTAGACCAAAAGCT.....TT 1992
      ::|||::: ||| ::|||:::
741 lGlyAlaLysGlyAlaThrSerIleTyrValLysAsnLeuThrValVal 758
      ::|||::: ||| ::|||:::
1993 AGGAAGTGAA.....GTTT 2006
      ::|||::: ||| ::|||:::
758 yAspGlyLysGluGluLysValGluPheAspLys 769
      ::|||::: ||| ::|||:::
2007 AGGATGCTTCTAAAGTCACCTACATCAGATAGATAA 2041
      ::|||::: ||| ::|||:::

```


OM of: US-09-754-947-1 to: N_Geneseq_1101.* out_format : pfs
Date: Mar 23, 2002 9:17 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODL=frame+g2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09754947/unat_22032002_125445_25485/app-query.fasta.1.854
-DB=N_Geneseq_1101 -OFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -YCAPOP=10.000 -YGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YCAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=trans40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09754947@cgn1_1.361 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPRX
-WAIT -THREADS=1

Search information block:

Query: US-09-754-947-1
Query length: 785
Database: N_Geneseq_1101.*
Database sequences: 930621
Database length: 428662619
Search time (sec): 173.620000

Score list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation	
/SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH28330			3875.00	4873.00	2.4e-263	2370	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH75488			419.50	517.17	9.9e-21	2766	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH71798			419.50	517.17	9.9e-21	2766	
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH22751			415.00	511.51	2.0e-20	2763	
/SID2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAO99440			373.00	454.98	2.9e-17	4197	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH75487			248.50	299.21	1.4e-08	3687	
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH22748			248.50	299.21	1.4e-08	3687	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH71797			248.50	299.02	1.4e-08	3687	
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH78246			248.50	299.02	1.4e-08	3687	
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH78245			248.50	299.88	1.9e-08	4888	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH78245			241.50	285.19	8.3e-08	6744	
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH71476			240.50	283.25	1.1e-07	7291	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH92499			238.50	280.78	1.5e-07	7253	
/SID2/gcgdata/geneseq/geneseqn/NA1991.DAT:AAO11789			231.00	271.45	4.8e-07	7152	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH65950			225.50	216.77	0.0005	183012	
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH42063			224.00	269.32	6.3e-07	3388	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52184			224.00	269.27	6.4e-07	3306	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH90993			221.50	262.91	1.4e-06	4803	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAO49511			218.50	259.31	2.3e-06	4702	
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:AAO72286			214.50	250.91	6.7e-06	6342	
/SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH59105			214.00	250.43	4.3e-06	4287	
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:AAO72285			213.00	252.41	5.7e-06	4794	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH90999			212.50	250.81	4.1e-06	3294	
/SID2/gcgdata/geneseq/geneseqn/NA1992.DAT:AAH24123			211.50	250.65	6.9e-06	4612	
/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH74374			211.50	245.75	1.3e-05	8155	
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH7593			209.00	250.17	7.4e-06	3945	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52178			209.00	251.42	6.3e-06	2928	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52177			208.00	249.22	8.3e-06	3263	
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAO49507			208.00	245.66	1.3e-05	4837	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH90995			208.00	245.66	1.3e-05	4840	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52197			207.00	244.50	1.0e-05	3443	
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAO49506			207.00	244.09	1.6e-05	5116	
/SID2/gcgdata/geneseq/geneseqn/NA1994.DAT:AAO72293			207.00	244.09	1.6e-05	5116	
/SID2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAH52195			206.00	242.83	1.9e-05	5116	

/SID2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAH90996	206.00	237.80	3.6e-05	9171
/SID2/gcgdata/geneseq/geneseqn/NA1993.DAT:AAO49508	206.00	237.76	3.6e-05	9220
/SID2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAH36122	205.00	242.00	2.1e-05	4865
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH20196	204.00	244.10	1.6e-05	3294
/SID2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAH20178	204.00	244.10	1.6e-05	3294

seq_name: /SID2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH28330

seq_documentation_block:

ID AAH28330 standard; DNA; 2370 BP.

AC AAH28330;

DT 05-SEP-2001 (first entry)

DE Nucleotide sequence of a bacterial surface array protein (SAP).

KW Surface array protein; SAP; bacterial detection; ss.

OS Bacillus anthracis.

XX Key Location/Qualifiers

FT CDS 1..2349

FT /product="surface array protein"

FT /note="the nucleotides encoding residues 1-2 of

FT the protein are not given"

PN WO200149823-A2.

PD 12-JUL-2001.

PF 04-JAN-2001; 2001WO-US00358.

PR 06-JAN-2000; 2000US-0174901.

PA (BIOS-) BIOSITE DIAGNOSTICS INC.

PI Lee BA, Flores BM, Valkirs GE;

WP: 2001-418358/44.

P-PSDB; AAB84651.

PT Novel methods and kits for detecting the presence of Bacillus anthracis

PT In a test sample -

XX Disclosure; Page 61-62; 62pp; English.

CC The present sequence encodes a bacterial surface array protein (SAP).

CC SAP is used in the method of the invention. The specification describes

CC a method for detecting the presence of Bacillus anthracis in a test

CC sample. The method comprises contacting the sample with a capture

CC reagent and detecting whether the a SAP is bound to the capture

CC reagent, which is indicative of the presence of B. anthracis in the

CC sample. The method is useful for detecting the presence or absence of

CC B. anthracis in a test sample.

XX Sequence 2370 BP; 940 A; 380 C; 459 G; 591 T; 0 other;

XX alignment_scores: Quality: 3875.00 Length: 783

XX Ratio: 4.949 Gaps: 0

XX Percent Similarity: 100.000 Percent Identity: 100.000

XX alignment_block: US-09-754-947-1 x AAH28330 ..

XX Align seg 1/1 to: AAH28330 from: 1 to: 2370

3 Lysrhphpeproaspvalproalaasphistrpolsyllleaspserrileas 19

1 AAACATTCCTCCAGCGTTCCTGCTGATCCTGCGGGAATTCATTCATTA 50

19 nrrytleuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheG 36
|||||
51 CTACTTAGTAGAAAAAGGCGCAAGTTAAAGGTAACGACAAAGAAATTTGG 100
36 LurProGlyLysGluLeuThrArgAlaGluAlaAlaThrMetCysLeuGln 52
|||||
101 ACCCTGGAAAAAGAAATTAAGTCTGCGAAGACGATCAATGATGGCTCAA 150
53 IleLeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSe 69
|||||
151 ACTTAACTTACCAATCGATTAAGATGCTAAACCATCTTCTGCTGACTC 200
69 rGlnGlyGlnTrpTyrThrProPheIleAlaAlaValGluLysAlaGlyV 86
|||||
201 TCAAGGCCAATGCTACACTCCATTCATTCGCACTGTAGAAAAAGCTGGCG 250
86 aLIleLysGlyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArg 102
|||||
251 TTAATTAAAGGTACAGGAAACGGCTTTGAGCCAAACGGAAAAATCGACCC 300
103 ValSerMetAlaSerLeuLeuValGluAlaTyrClyLysLeuAspThrLysVa 119
|||||
301 GTTCTATGCGCATCTCTTCTTGTAGACGCTTCAAAATTTAGATCTAAAGT 350
119 LAsnGlyThrProAlaThrLysPheLysAspLeuGluThrLeuAsnTrpG 136
|||||
351 AAACGGTACTCCAGCAACAATTAATCAAGATTAGAAACATTAACTGGCG 400
136 LylsGluLysAlaAsnIleLeuValGluLeuGlyIleSerValGlyThr 152
|||||
401 GTAAAGAAAAAGCTTAACATCTTAGTTGAATTAGAAATCTCTGTGGTACT 450
153 GlyAspGlnTrpGluProLysLysThrValThrLysAlaGluAlaAlaG 169
|||||
451 GGTGATCATGGGAGCGCTAAGAAACTGTAACTAAGCGAAGCGTCTCA 500
169 nPheIleAlaLysThrAspLysGlnPheGlyThrCysAlaAlaLysValG 186
|||||
501 ATTCAATGCTTAAGACTGACAMCGACGTGCTGACAGACGCAAAAGTTG 550
186 LuSerAlaLysAlaValThrThrGlnLysValGluValLysPheSerLys 202
|||||
551 AATTCGCAAAAGCTTTACAACTCAAAAAGTAGAAGTTAAATTCAGCAAA 600
203 AlaValGluLysLeuThrLysGluAspIleLysValThrAsnLysAlaAs 219
|||||
601 GCTGTTGA AAAATTAAGTAAGACATATCAAAAGTAACTAACAAGCTAA 650
219 nAsnAspLysValLeuValLysGluValThrLeuSerGluAspLysArgS 236
|||||
651 CAACGATTAAGTACTAGTTAAAGAGCTTAACCTTATCAGAAAGATAAAGAT 700
236 GrrAlaThrValGluLeuTyrSerAsnLeuAlaAlaLysGlnThrCyrThr 252
|||||
701 CTGCTACAGCTTGAATTAATTAAGTAACTTAGCGCTAAACCAAACTTACACT 750
253 ValAspValAsnLysValGlyLysThrGluValAlaValGlySerLeuG 269
|||||
751 GTAGATGTAACCAAGCTTGTAAACACAGAGCTGTAGCTTCTTTAGA 800
269 uAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspGluProT 286
|||||
801 AGCAAAAACCAATCGAAATGCTGACCAAAACAGTTGTAGCTGATGAGCCAA 850
286 hrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValValSer 302
|||||
851 CAGCAATTCACATTCACAGTTAAAGATGAAACGGTACTGAAAGTTGTTTCA 900
303 ProGluGlyIleGluPheValThrProAlaAlaGluLysIleAsnAlaL 319
|||||
901 CCAGAGCGCTATTGAATTTGTAACGCGCACCTGCAGAAAAAATTAATGCAAA 950

319 sGlyGluIleThrLeuAlaLysGlyThrSerThrThrValLysAlaValTr 336
|||||
951 AGGTGAATACACTTTAGCAAAAAGTACTCACTCACTGTAATAAGCTGTTT 1000
336 YrLysLysAspGlyLysValValAlaGluSerLysGluValLysValSer 352
|||||
1001 ATAAAAAAGACGCTAAAGTACAGTCAAGATTAAGAAAGTAAAGTTTCT 1050
353 AlaGluGlyAlaAlaValAlaSerIleSerAsnTrpThrValAlaGluG 369
|||||
1051 GCTGAAGGTGCTGCACTACCTTCAATCTTCACTACGACGTTGCGAACA 1100
369 nAsnLysAlaAspPheThrSerLysAspPheLysGlnAsnAsnLysValTr 386
|||||
1101 AAATTAAGCTGACTTACTTCTTAAAGATTTCAAACAAAAACATTAAGTTT 1150
386 YrGluGlyAspAsnAlaTyrValGlnValGluLeuLysAspGlnPheAsn 402
|||||
1151 ACGAAGGCGCACACGCTTACGTTCAAGTAGAATTAAGATCAATTTTAC 1200
403 AlaValThrThrGlyLysValGluTyrGluSerLeuAsnThrGluValAl 419
|||||
1201 GCAGTAACACTGGAAGAGTTGAATATGAGTCGTTAAACACAGAAAGTTGC 1250
419 aValValAspLysAlaThrGlyLysValThrValLeuSerAlaGlyLysA 436
|||||
1251 TGTAGTAGTAAGCTACTGGTAAGTAGTACTGATTAATCTGCAGGAAAG 1300
436 LProValLysValThrValLysAspSerLysGlyLysAlaLeuValSer 452
|||||
1301 CACCGTAAGAAAGTAACTGTAAAGATTCAAAAGGTAAAGCACTTGTTC 1350
453 HisThrValGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleL 469
|||||
1351 CACACAGTTGAATTTGAAGCTTCCGCTCAAAAAGCAATGAAGAACATTTAA 1400
469 sLeuGluLysThrAsnValAlaLeuSerThrLysAspValThrAspLeuL 486
|||||
1401 ATTAGAAAAAAGCTAAATGTAAGCGCTTTCTTACAAAAGATGTACAGATTAA 1450
486 YsValLysAlaProValLeuAspGlnTyrGlyGluPheThrAlaPro 502
|||||
1451 AAGTAAGAGCTCCAGTACTAGTCAATACGGAAGAGCTTTACACCTCT 1500
503 ValThrValLysValLeuAspLysAspGlyLysGluLeuLysGluGlnL 519
|||||
1501 GTAACAGTGAAGTACTGATTAAGATGTAAGATTAAGATTAAGAACAAAA 1550
519 sLeuGluAlaLysTyrValAsnArgGluLeuValLeuAsnAlaAlaGlyG 536
|||||
1551 ATTAGAACCTAAATATGTGAACACAGAAATTGATTCTGAATGCACAGGTC 1600
536 InGluAlaGlyAsnTrpThrValValLeuThrAlaLysSerGlyGluLys 552
|||||
1601 AAGAGCTGCTGTAATTAACAGTTGATTAACTGCAAAATCGGTGTA AAAA 1650
553 GluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSe 569
|||||
1651 GAAGCAAAAAGTACTATAGCTCTGAATTAAGATTCACAGGTGCACTTCTC 1700
569 rLysPheGluValArgGlyLeuAspThrGluLeuAspLysTyrValThrG 586
|||||
1701 TAAATTTGAAGTTCGTGCTTGAACACAGAAATTGATTAATATGTACTG 1750
586 LgluAsnGlnLysAsnAlaMetThrValIserValLeuProValAspAla 602
|||||
1751 AGGAAACCAAAAGAAATGCAATGACTGTTTCAAGTTCTTCCGTGATGATCA 1800
603 AsnGlyLeuValLeuLysGlyAlaGluAlaAlaGluLeuLysValThrTh 619
|||||
1801 AATGATTAGTATTAAGTAAGTGCAGAAAGCACACTGACATTAAGATACAC 1850
619 rThrAsnLysGluGlyLysGluValAspAlaThrAspAlaGlnValThrV 636

```

|||||
1851 AACAAACAAAGAGTAAGAGTAGAGCAGACGATGCACAGCTTACTG 1900
    alGlnAsnAsnSerValIleThrValGlyGlnGlyAlaValAlaGlyGlu 652
636 |||||||
    TACAAATTAACAGTGTATTACTGTGTGTCAGAGGCAAAAGCTGTAG 1950
653 ThrTyrLysValIleThrValLeuAspGlyLysLeuIleThrThrHisSe 669
    |||||||
1951 ACTTATTAAGTAACAGTGTACTAGATGTGTAATTAATCACAACTCATTC 2000
669 rPhelValValAspThrAlaProThrAlaLysGlyLeuAlaValGluP 686
    |||||||
2001 ATTCAAAAGTTGTGATACGACCAACACTGTAAAGGATTACAGATAGAT 2050
686 herThrSerThrSerLeuLysGlyValAlaProAsnAlaAspLeuLysAla 702
2051 TTACACACACATCTCTTAAAGAGTAGCTCCAAATGCGATTAAAGCT 2100
703 AlaLeuLeuAsnIleLeuSerValAspGlyValProAlaThrThrAlaLy 719
2101 GCACCTTTAAATATCTTATCTGTGATGCTGCTACCTGCAGCTACAGCMA 2150
719 sAlaThrAlaSerAsnValGluPheValSerAlaAspThrAsnValAla 736
2151 AGCAACAGCTTCTAATGTAGAAATTTGTTCTGCTGACCAAAATGTTGTAG 2200
736 IagIAsnGlyThrValGlyAlaLysGlyAlaIleThrSerIleTyrValLys 752
2201 CTGAAAATGCTACAGTGTGTCGCAAAAGTGCACATCATCTATCTGTA 2250
753 AsnLeuThrValValLysAspGlyLysGluGlnLysValGluPheAspLy 769
2251 AACCTGACAGCTGTGTAAGATGCAAAAGCAAAAGTAAGAATTGTGATA 2300
769 sAlaValGluValAlaValSerIleLysGlyAlaLysProAlaThrLys 785
2301 AGCTGTACAAAGTGTGACGTTCTATTAAAGAAAGCAAAACCTGCACAAA 2349
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT.AAT75488
seq_documentation_block:
ID AAT75488 standard; DNA: 2766 BP.
XX
AC AAT75488;
XX
DT 27-FEB-1998 (first entry)
XX
DE DNA for Bacillus stearothermophilus S-layer protein sbs-B.
XX
KW S-layer; sbs-B; vaccine; adjuvant; carrier; hybridisation assay;
KM molecular spinning nozzle; molecular laser; ds.
XX
OS Bacillus stearothermophilus.
XX
FH key Location/Qualifiers
FT CDS 1..2766
FT sig_peptide 1..93
FT mat_peptide 94..2763
FT /*tag- b
FT /*tag- c
XX
DE19603649-A1.
XX
PD 07-AUG-1997.
XX
PF 01-FEB-1996; 96DE-1003649.
XX
PR 01-FEB-1996; 96DE-1003649.
XX
PA (LUBI/) LUBITZ W.
PA (SLEY/) SLEYTR U.

```

```

XX
PI Kuen B, Lubitz W, Sleytr U;
XX
DR MPI; 1997-394558/37.
XX
PT P-PSDB; AAM22863.
XX
PT Preparation of S-layer proteins by expressing sbs-A gene in Gram
PT negative bacterium - or new sbs-B gene in any host, also new
PT recombinant proteins containing heterologous inserts, e.g.
PT epitope(s), useful as vaccines and adjuvants
XX
PS Claim 26; Pages 19-23; 31pp; German.
XX
CC The present sequence encodes the Bacillus stearothermophilus PV72
CC S-layer protein, sbs-B. S-layer structures can be used as vaccines
CC or adjuvants, particularly when they include a bacterial ghost that
CC may contain additional epitopes in its membrane. Other uses of
CC recombinant sbs-B, depending on the nature of the inserted peptide,
CC are as an universal carrier for biotinylated reactants for use in
CC immunological or hybridisation assays (the insert is streptavidin),
CC to induce immune responses (epitopes), as a reagent for removing
CC cytokine or toxin from serum (antigenic epitopes), as a molecular
CC laser (luciferase).
XX
SQ Sequence 2766 BP; 966 A; 465 C; 534 G; 771 T; 0 other;

alignment_scores:
Quality: 419.50 Length: 879
Ratio: 0.891 Gaps: 40
Percent Similarity: 53.584 Percent Identity: 24.005

alignment_block:
US-09-754-947-1 x AAT75488 ..

Align seg 1/1 to: AAT75488 from: 1 to: 2766

4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
:::|::| |::|::| ::::: |::|::|::|::|
97 AGCTTCACAGATGTTGCCGCCAATAT.....AAGATGCGATGATTT 140
20 rLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluP 37
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
141 CTTAGTATCAACTGTGTGCACAAAGGTAAACAGAAACAAATTCGGCG 190
37 roGlyLysGluLeuThrArgIaGluAlaIleThrMetMetAlaGlnIle 58
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
191 TTTACGATGAAATCATCTGCTAGTCGCGCAGTTATTCTTGCAAGATA 240
54 LeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSergI 70
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
241 TTAAAACTAGACGTTGACACGCAAAAGACGACGCTTCACAGATGTC 290
70 nGlyGlnTrpTyrThrProPheIleAlaValAlaGluLysAlaGlyValI 87
::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
291 AAAAGAC...CGTGCAAATATACGTCACGCGCTGTGAACCTGCGGTAT 337
87 IeLysGlyThrGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
338 TAAACGTTAAAGCACCTGGCAATTTGGTGCATACGACCCATTAACTGCG 367
103 ValSerMetAlaSerLeuLeuValGluAlaTyrLysLeuAsp..... 116
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
388 GTTGAATATGCAAAATATCATCGCAACCGTTAAATTAAGCTGACGA 437
117 .....ThrLysValAsnGlyThrProAlaThrLysPheI 128
438 TGTAAAACTTCATTCATCTGATGTAAACGATACACAGGACCATACGTA 487
128 ysAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
488 AAGCGCTT.....TATAAATACGAAGTAAACCAAAAGTTTAAA 525

```

[illegible]

411 tTgCISerLeuAsnThGluValAlaValAlaSprsAlaThnGly 427
1351 GTCAGTCTTTTAAATCCAAATTATGCAACAGCGCTTAATGTAAGCA 1400
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
1401 GCTCCTTGTCACAGCAATGCTGGCCAATCTGGAAAACCTTCATTGAG 1450
440 aThrValLysAspSerLysGlyLysAlaLeuValSerHisThrValGlu 456
1451 TAACATTAAAGATAT.....ACAAAAAGAACTTTACA 1485
457 TiegValaPhaAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473
1486 GTTGATGTAAAAAAGACCCTGTATTACAGATATAAAGTAGATGCAC 1535
473 rAsnValAlaLeuSer.....ThrLysAsp 482
1536 TTCTGTTAACTTTCGATGAAGCTGTGGCGGGGGAACCTGAAGAG 1585
482 aThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyLysGlu 498
1586 TTAACCAAAAAACGATTAAAGTAAGTGCAAGTTAGACCAATTACGTAAAGA 1635
499 phe.....ThrAlaPr 502
1636 ATTAATATTTGTACAAAAGTAAGTACTGTTCACAACTATACAGAAAG 1685
502 oValThrValLysValLeuAspLysAsp.....G 512
1686 ACTAGTATTATAAATGTAAATAGCGATATAATACATTACTTGTATAGCG 1735
512 LysLysGluLeuLysGluGln..... 518
1736 GCATATAGTGCACACTGACCAATTTGTTGCTGTGCACAAAAGACAAAT 1785
519LysLeuGluAlaLysTyrValAlaSerGluLeuValLeuAs 532
1786 GTCAATGTAAGTAAAGTGAAGTTAAATATTTCCAAAT..... 1821
532 nAlaAlaGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLys 549
1822 ...GCTAGTGCACACACCACTCACTCACTAAACATTTACTGTTATG 1867
549 eGlyGluLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1868 TAGTAAATGTAAAGCTGACGCTACACACAGTAGATTAGATTGTGACA 1917
565 ProGlyAlaPheSerLysPheGluValAlaGlyLysAspThrGluLeuAs 581
1918 CCT.....TCATAATTTGATGTAAATCTCCAAACACT..... 1950
581 pLysTyrValThrGluGluAsnGlnLysAsnAlaMetThrValSerVal 598
1951 .GCTTCTACGACAGTGTGATTATTATAATTTGGAAAGTTCAGACTTT 1999
598 euProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaAlaGlu 614
2000 ACACACTCGATTCAAATGCT.....AGACGTCAA 2028
615 LeuLysValThrThrThr.....AsnLysGluGly 625
2029 AAAAAAGTTACTCCCAAGTCAACTGACCTGTAGCTCAAAAAA 2078
625 sGluValAspAlaThrAspAlaGlnValThrValGlnAsnAsnSerVal 642
2079 AAAAGTTAATGGAATGTATTACA...TTCAAGGGACGCAAGAT 2122
642 LeuThrValGluGlnGlyAlaLysAlaGluGlnThrTyrLysValThrVal 658
2123 TAACGCTATCAACTTCTTAGTACAGGAAC..... 2154
659 ValLeuAspGlyLysLeuIleThrThnHisSerPheLysValAlaSprh 675

```
2155 ...GTAATGCA..... 2163
XX
XX
675 rAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeuL 692
2164 .....ACAGCAGAGGAATGACA.....A 2182
692 ysgLValAlaProAsnAlaSplLeuAlaValLeuAsnIleLeu 708
2183 AACGTAT...CCAGGGAATATATCAACTGCA..... 2214
709 SerAlaSplGlyValProAlaThrThrAlaLysAlaThrAlaSer..As 724
2215 .....AGGTACTGCCAGTGCACACAGTACCAACAGTCTGTTAC 2255
724 nValGluPheValSerAlaAspThrAsnValAlaValGluAsnGlyThr 741
2256 TGTAACTTAATTCAGAGTGTATGATTGATTGACATTGAGAAATTATAT 2305
741 aLGIAlaLysGlyAlaThrSerIleTyValLysAsn..... 753
2306 TCGGTGAATTCAGCCCTACCAATTA...GTCAAGATGAAGACATCAAC 2352
754 .....LeuThrValValLysAspGlyLysGluGlnLysValGluPheAs 768
2353 GAATTTATTCAGTTCAAAAGCGCTAAATATGATATTGTTGTTAA 2402
768 pLysAlaValAlaGlnValAlaValSerIleLysGluAla 780
2403 TAAACCGCTT.....GTACGGTTAAAGATGCA 2430

seq_name: /SIDS7/gcgdata/geneseq/geneseq/NA2000.DAT.AAA71798
seq_documentation_block:
ID AAA71798 standard; DNA; 2766 BP.
XX
XX AAA71798;
XX
XX 08-JAN-2001 (first entry)
XX
XX B. stearothermophilus sbsb DNA.
XX
XX S-layer gene; sbsb: antibacterial; vaccine; adjuvant; bioreactor;
XX polyhydroxyalkanoate synthesis; ds.
XX
XX Bacillus stearothermophilus.
XX
XX Key Location/Qualifiers
XX CDS 1..2766
XX FT /tag= a
XX FT /product= "sbsb"
XX FT sig_peptide 1..93
XX FT /tag= b
XX FT mat_peptide 94..2763
XX FT /tag= c
XX
XX DE19903345-A1.
XX
XX 03-AUG-2000.
XX
XX 28-JAN-1999; 99DE-1003345.
XX
XX 28-JAN-1999; 99DE-1003345.
XX
XX (LUBI/) LUBITZ W.
XX
XX Lubitz W;
XX
XX WPI: 2000-533868/49.
XX
XX P-PSDB: AAB10626.
XX
XX Host cell, useful e.g. as bioreactor for production of
XX poly(hydroxyalkanoate), containing two or more recombinant
PT
```

```
PT polypeptides, with at least one in carrier-bound form
XX
XX Disclosure; Page 16-19; 26pp; German.
XX
XX This invention describes a novel host cell (A) comprising at least two
XX functional recombinant polypeptides (I), at least one being in carrier
XX bound form. The products of the invention have antibacterial activity.
XX (A), or, where bacterial, their ghosts (B), are useful as vaccines or
XX adjuvants (specifically for presentation of immunogenic epitopes of
XX pathogens or analogous immunostimulatory polypeptides, e.g. cytokines),
XX or preferably, as enzyme reactors for performing a cascade of reactions,
XX specifically synthesis of poly(hydroxyalkanoate). Localization of
XX individual (I), specifically enzymes, in separate cellular compartments
XX avoids adverse reactions between products and substrates, when being used
XX as bioreactors. (I) can be produced in carrier-bound form without loss of
XX function. This sequence encodes the Bacillus stearothermophilus S-layer
XX protein sbsb which is used to illustrate the method of the invention.
XX
XX Sequence 2766 BP; 996 A; 465 C; 534 G; 771 T; 0 other;
SQ

alignment_scores:
Quality: 419.50 Length: 879
Ratio: 0.891 Gaps: 40
Percent Similarity: 53.584 Percent Identity: 24.005

alignment_block:
US-09-754-947-1 x AAA71798 ..

Align seg 1/1 to: AAA71798 from: 1 to: 2766

4 ThrPheProAspValProAlaAspHisTrpGlyIleAspSerIleAsnTy 20
:::|::| |::|::| ::::: |::|::|::|::|
97 AGCTTCACAGATGTTGCCGCCAATAT.....AAAGATCGATCGATT 140
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
20 rLeuValGluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGlu 37
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
141 CTGATATCAACTGTGTCACAAAGGTAAGACAAACAAATTCGGCG 190
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
37 roGlyLysGluLeuThrArgLacLysAlaValAlaThrMetAlaGlnIle 53
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
191 TTTACGATGAATATCATCTGCTAGATCGCGCATATTCTTCAAGATA 240
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
54 LeuAsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerG 70
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
241 TTTAAACTAGACGTTGACACGCAAGACGCGCTTCACGATGTCGCC 290
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
70 nGlyGlnTrpTyThrProPheIleAlaValAlaGluLysLacIleValI 87
::: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
291 AAAAGAC...CGTCAAAATATCGTCAACGCGCTTGTAAGACGTGGCAT 387
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
87 lElysglyThrGlyAsnGly...PheGluProAsnGlyLysIleAspArg 102
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
338 TAAACGGTAAAGCACCTGCAAAATTGGTGCTACGACCCATTAAACGCC 387
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
103 ValSerMetAlaSerLeuLeuValGluAlaTyrlsLeuAsp..... 116
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
388 GTTGAATATGCGCAAAATATCGCGATCAAAATTAAGAGCTGACGA 437
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
117 .....ThrLysValAsnGlyThrProAlaThrLysPheL 128
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
438 TGTAAACCTTCATTCACTGATGAACGATACATGGCCACCATACGTAA 487
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
128 yAspLeuGluThrLeuAsnTrpGlyLysGlyAlaAsnIleLeuVal 144
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
488 AAGCGCTT.....TATAATACGAATACCAAAAGGTTAAA 525
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluPolysLysTh 161
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
526 CACCACAGAGCTCGTGCATACC.....AAAAA 554
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
161 rValThrLysLacLysAlaGlnPheIleAlaLysThrAspLysGlnP 178
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```


488 AACGCTTTATTAATACGAAGTAACAAAGCTAA..... 522
145 GluLeuGlyIleSerValGlyThrGlyAspGlnTrpGluProLysIleTh 161
523ACACCACACAGCTTGGTGCAATACCAAA 551
161 rValThrLysAlaGluAlaGlnIleIleAlaLysThrAspLysGlnP 178
552 CACACACCGCGGTGACTTGGCCAAATTTGTATATACAGCGGTCAATATT. 600
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrTrgIn 194
601AATGACAGTCCAGAAATATGTTGAAGTAACCTCGGTTAATGCTACT 645
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGlns 211
646 ACAGTGAACCTAACATTCAATACGCAATTT.....GCTGA 680
211 pIleLysValThrAsnLysAlaAsnAspLysValLeuValLysGluV 228
681 TGTGATTTCACAAATTTTGTCTATCGATTAACGGTTAACTGTTACTAAAG 730
228 aLThrLeuSerGluAspLysArgSerAlaThrValGluLeuTrpSerAsn 244
731 CAACCTTTCTCGTGAATAAAATCCGTAGAGGTGTGGTAATAATAACCG 780
245 LeuAlaAlaLysGlnThrTyThrValAspValAsnLysValGly..... 259
781 TTTACTCGTATACGGAATATACAAATTACAGCAGCAGCATTAATAATT 830
260 LysThrGluValAla.....ValGlySerLeu. 268
831 AAAAGCGGACCGCTAAGCAATTAAGTGAATTTGTTGGTCTGCTTC 880
269 ..GluAlaLysThrIleGluMetAlaAspGlnThrValAlaAlaAspGlu 284
881 AAGATCCGTAAGTGTTCACATAAATATAGTTCGCTTAAGTTGGAGAG 930
285 ProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluValVa 301
931 GAATCTGTGTTTA.....ACTGTAAGAATCAGGATGCCAAGAATGTGT 974
301 L.....SerProGluGlyIleGluPheValT 310
975 AGGTGCTAAGTAGAAGTAAGTACTTCTTAATCTAATATGTTGTGTGTTT 1024
310 hProAlaIleGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLys 326
1025 CAAGTGGCCAGATATCAGTATCTGCTGCTAAAGTTACAGCTGTAACCG 1074
327 GlyThrSerThrThrValLysAlaValTyLysLysAspGlyLysValVa 343
1075 GGAAAGCATGATGTACTGCAAAATTTACATTACCGATGCGTGTGTTACT 1124
343 AlaGluSer.....LysGluValLysValSerAlaGlu 355
1125 AACCAATTAATTAAGTAGACAGTTACAGAGGCTGTGCAAGTACAA 1174
355 ValAlaAlaValAlaSerIleSerAsnTrpThrValAlaGluGln..... 369
1175 ATCAAGGATTTACTTATGATTAATCTTCTATATGCTCCACAGAAATACA 1224
370AsnLysAlaAspPheThrSerLysAspPheLysGluAsnAs 383
1225 GTTGATTTAAACAAGCTGAGAAAGTAACCTCAATGTTGCTGGAGAAAC 1274
383 nLys.....ValTyGlu.....GlyAspAsnAlaTyValG 394
1275 TAAAAACAGTTGCAATGTATGATACTAAAAACGCTGATCCTGAAACTAAC 1324
394 lValGluLeuLysAspGlnPheAsnAlaValThrGlyLysValGlu 410
1325 CTGTTGATTTCAAGAT.....GCAACT 1347

411 TyrgIleSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLy 427
1348 GATAGCTCATTAATCCATTTATTCGAACAGCTGCTATTAAATGAGTGA 1397
427 sValThrValLeuSer.....AlaGlyLysAlaProValLysV 440
1398 GCTCCTTCACACAGCTAATGCTGGCCAAATCTGGAAGAGCTTCATTGGA 1447
440 aLThrValLysAspSerLysGlyLysAlaLeuValSerIleThrValGlu 456
1448 TACATTTTAAAGATAT.....ACAAAGAACATTTTCA 1482
457 lIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGluLysTh 473
1483 GTTGATGTGAAAAAAGACCCCTGATTACAGATTTAAAGTGAATGCAAC 1532
473 rAsnValAlaLeuSer.....ThrLysAspV 482
1533 TTTCTGTTAACTTTCGATGAAGCTTTGCCGCGGAGGAAGTTGAAGAG 1582
482 aLThrAspLeuLysValLysAlaProValLeuAspGlnTyrgLysGlu 498
1583 TTACCAAAAAAAGATTAAGAAGTGCAGTGCACCAATACGGTAAGAA 1632
499 Phe.....ThrAlaPr 502
1633 ATTAATTTGTTACAAAAAGCTAAGTTACTGTTACACTAATATACGAAGG 1682
502 oValThrValLysValLeuAspLysAsp.....G 512
1683 ACTAGTTATTAATAAATGTAAATAGCATTAATCAATTGACTTTGATAGC 1732
512 TyLysGluLeuLysGluGln..... 518
1733 GCAATATGTCACCACTGACCAATTGTTGCTGCTGCACAAAGACAAAT 1782
519LysLeuGluAlaLysTyThrValAsnArgGluLeuValLeuAs 532
1783 GTCAATGTGAAGTAGAAGTTAATATTTCAAAAT..... 1818
532 nAlaAlaGlyGlnGluAlaGlyAsnTyThrValValLeuThrAlaLys 549
1819 ...GCTAGTGACACACACCACTTCACTAAACATTAAGTGTGTAATG 1864
549 eGrgLyuLysGluAlaLysAlaThr...LeuAlaLeuGluLeuLysAla 564
1865 TAGTGAAATTAAGTACGCTACACACGATGAGTTAGATTTGTACGA 1914
565 ProGlyAlaPheSerLysPheGluValArgGlyLeuAspThrGluLeuAs 581
1915 CCT.....TCTGAATATGATGTGAATGCTCCAAACACT..... 1947
581 pLysTyThrValThrGluGluAsnGlnLysAsnAlaMetThrValSerVal 598
1948 ..GCTTCTCTGACAGATGTGATTTATTAATTTGCAAGAGTGTGATTT 1996
598 euProValAspAlaAsnGlyLeuValLeuLysGlyAlaGluAlaLeuGlu 614
1997 ATACACTGATTTCTATAGTAAACGCTCTTAATAAAAGTACTACCACTGCA 2046
615 LeuLysValThrThrAsnLysGluGlyLysGluValAspAlaThrAs 631
2047 ACTACACTGTGTAGTAAAT.....GATTAATGTTGAAGTAAATGCG 2087
631 pAlaGlnValThrValGlnAsnAsnSerValIleThrValGlyGlnGly 648
2088 GAATGATTTTACAAATTCAGGGAACGATGAATTAACGCTA.....T 2128
648 lAlaLysAlaGlyGluThrTyTrpLysValThrValValLeuAspLysLeu 664
2129 TAACCTTCTTCTAGTACAGTAACGTTGATGTACAGCTGATGGA..... 2172

```

665 Ilerthrhis...SerphelyValValAspThrAlaProThrAlaIly 680
      |||||   :   |||   :   :   :   :   :   :   :
2173 ATTACAAACCTATTCCAGTAAATATATCAACTCGAAGTACCTGC 2222
      :   :   :   :   :   :   :   :   :   :   :
680 sGlyLeuAlaValGluPheThrSerThrSerLeuSgluValAlaProA 697
      :   :   :   :   :   :   :   :   :   :   :
2223 CAGTGCACACGTACACACAGTCTGTACTGTAAAGCTTAATTCACGTG 2272
      :   :   :   :   :   :   :   :   :   :   :
697 snAlaAspLeuLysAlaAlaLeuLeuSnIleLeuSerValSglYal 713
      :   :   :   :   :   :   :   :   :   :   :
2273 ATATGATTTACATTGAGATTTAATATTCGGTAAATTGAC..... 2316
      :   :   :   :   :   :   :   :   :   :   :
714 ProIathrThrAlaLysAlaIleThrAlaSerAsnValGluPheValSerAl 730
      :   :   :   :   :   :   :   :   :   :   :
2317 CCTACACATTTAGTCAAGATGACATCAAC...GAATTTATTTGCAGT 2363
      :   :   :   :   :   :   :   :   :   :   :
730 aspPThrAsnValValAlaGluAsnGlyThrValGlyValAlaSglYalA 747
      :   :   :   :   :   :   :   :   :   :   :
2364 T.....TCMAAACGGCTAAATAATGATG 2386

747 hrSerIleTyValLysAsnLeuThrValValLysAsp 759
      :   :   :   :   :   :   :   :   :   :   :
2387 GATATTGTATTAATAAACCGCTGTGAAGCTTAAGAT 2424

seq_name: /STD2/gcgdata/geneseq/geneseq/NA1995.DAT:AA099430
seq_documentation_block:
ID   AA099430 standard; DNA; 4197 BP.
XX
XX   AA099430:
XX
XX   22-DEC-1995 (first entry)
XX
XX   B. sphaericus SLP gene.
XX
XX   Surface layer protein; SLP; fusion protein; vaccine; antigen;
KM   surface expression; epitope; ds.
XX
XX   Bacillus sphaericus.
XX
XX   Key   Location/Qualifiers
FH   RBS   79..85
FT       /*tag= a
FT       95..3653
FT       /*tag= b
FT       95..184
FT       /*tag= c
FT       185..3850
FT       mat_peptide
FT       /*tag= d
XX
XX   WO9519371-A2.
XX
XX   20-JUL-1995.
XX
XX   13-JAN-1995; 95MO-EP00147.
XX
XX   14-JAN-1994; 94GB-0000650.
XX
XX   (SOLV ) SOLVAY SA.
XX
XX   Deblaere RY, Desomer J, Dhasee P;
XX
XX   WPI: 1995-263827/34.
XX
XX   P-PDB; AAR80530.
XX
XX   Host cell expressing surface layer protein fusion protein - used for
PT   host presentation of antigens and vaccine prodn.
XX
XX   Disclosure; Fig.6; 95pp; English.
XX
XX   A probe based on the N-terminal sequence of B. sphaericus P-1 (LMG
CC   P-13855) surface layer protein was used to screen a HindIII-
CC   generated library to isolate the slp gene. Promoter regions

```

```

CC   of the gene are used in genetic constructs providing surface
CC   expression of heterologous proteins in P-1 hosts.
XX
XX   Sequence 4197 BP; 1470 A; 731 C; 763 G; 1233 T; 0 other:
SQ
alignment_scores:
      Quality: 373.00      Length: 1009
      Ratio: 0.740      Gaps: 46
      Percent Similarity: 49.950      Percent Identity: 22.398
alignment_block:
US-09-754-947-1 x AA099430 ..
Align seg 1/1 to: AA099430 from: 1 to: 4197

16  AsperilleAsnTyrlLeuValGluLysGlyValAlaValSglYalAsnAspLys 32
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
227  GAACGAGTTCAAGCTTTAGTTGACCAAGCGCTAATCCAAAGGTGATACTAA 276
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
32  sGlyMetPheGluProGlyLysGluLeuThrArgAlaGluAlaIleThrM 49
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
277  CGGGAAGTTCACCCACTTAACACAGTAACGTCGTCACAAAGCTGCAGAA 326
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
49  eMetAlaGluIleLeuAsnLeuProIleAspLysAspAlaLysProSer 65
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
327  TCTTCACAAAAGCTTTAGATTTAGACCTAACGCGATGTA.....AAC 370
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
66  PheAlaAsp...SerGlnGlyIleTyrThrProPheIleAlaIleAla 81
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
371  TTCAAAGAGCTGAAGAGCTGGCTTGCTGACTACAACTCAATCGCTGCTGT 420
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
81  GluLysAlaGlyValIleLysGly...ThrGlyAsnGlyPheGluProA 97
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
421  TGTAGCTAACGCGATTTTGAAGGTGTTAGTGAACGATGATTCACCAA 470
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
97  snGlyLysIleAspArgValSerMetAlaSerLeuLeuValGluAlaTyr 113
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
471  ACAATCTTTTAACTGCTTGAAGCTGCTAAATAATTTAGTGAAGCATTC 520
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
114  LysLeuAspThrLysValAsnGlyThrProIathrLysPheLysAspLe 180
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
521  GGTTTAGAGGTGAAGCAGAT.....CTTAGCGAATTTGCTGACGC 561
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
130  uGluThrLeuAsn...TyrGlyLysGluLysAlaSnIleLeuValGluL 146
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
562  TTCACAGTAACCTGGCGCTAAATAATACTAGAAATCCAGTAGCTA 611
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
146  euGlyIleSerValGlyThr...GlyAspGlnTyrGluProLysLysThr 161
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
612  ACGGCATTTTCGAAGGTACTGATGCAACAACTTAACCTTAACACATCA 661
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
162  ValThrLysAlaGluAlaIleAlaGlnPheIleAlaLysThr...AspLysG 177
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
662  ATCACTGCTCAAGACTTTGACACTGCTTCAACGTRACAGTGCACAAAGT 711
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
177  nPheGlyThrGluAlaIleAlaLysValGluSerAlaLysAlaValThrThrg 194
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
712  TGAAGGTGAACCTCCAGAGAAGACGACATTTGTTAAAGCTTACAAACA 761
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
194  InLysValGluValLysPheSerLysAlaValGluLysLeuThrLys... 209
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
762  CAACTGTTGAAGTAACATTCGAGAGAAGAGTACTAACGTTCAAGCACTT 811
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
210  .....GluAspIleLysValThrAsn.....LysAlaAs 219
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
812  AACTTCAAAATCGAAGGTTTAGAATTAATAATCTTCTGTAAACAAC 861
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
219  nasnAspLysValLeuVal.....LysG 227
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
862  AAACAAAAGATTGTTGATTAAGTACTGAAGCTCAAAAGACGTGATAAG 911
      :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
227  luValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSer 243

```

```

912 AGTATGTTTAACTCTTGAC.....GGCAACAACATCGTGCTTAA 955
244 AsnLeuAlaIalysGlnThrTyrThrValAspValAsnLysValGly 260
956 GGGTGGCTGCT.....GTAGTCCCACTAAAGTTGACT 990
260 sThrGluValAlaVal..GlySerLeuGluAlaYsThrIleGluMet 276
991 AGTATCTCTCAGTTCAGGTAACCTGTCACAGAAAGTAAAGTTCAAG 1040
276 LaspGlnThrValValAlaLaspGluProThrAla.....LeuGln 289
1041 CTAAAGTAAGTCTGTCTGAAGGTCAATCTAAAGCTGTATTCCTGTACT 1090
290 PheThrValLysAspGlnLysnGlyThrGluValSer..... 302
1091 TTCACGTGACAGGTACACAAATGATGGCGTTGTACCAACTTAACAG 1140
302 ..... 302
1141 TGAAGCTTTAACAACGAGAGGTATCGCAACATCTCTACACTGTT 1190
303 ..ProGluGlyIleGluPheValThrProAlaAla..... 313
1191 ATAAAGAACGTACTGATGAAGTAAGTCTTATGCAACGTGATCGTCT 1240
314 .....GluLysIleAsnAlaLys 319
1241 AATTCTCACTGGTATGATATTCTGGGGTGTAGATACAAATCTTCACT 1290
319 sGlyGluIleThr.....LeuAlaYsGlyThrSerThrThv 332
1291 TGAAGAACTACTACAGTGTCTTCAATATATGTCGCAACAAACTT 1340
332 sLysValValTyrLys.....LysAspGlyLysVal..... 342
1341 ACAAAGTCTTATTAACCCCTAAACCTGTAAACGAGAACCAACAA 1390
343 .....ValAlaGluSerLysG 348
1391 ACATTTAATGTTGGTTTGTAGAAAACATGATGTTACTCTGATTAAC 1440
348 uValLysValSerAlaGluGlyAlaAlaValAlaSerIleSerAsnTPT 365
1441 ACCAAATGCTACAGTTAATGGCGTAAAGCATTAACATTAACCAATG 1490
365 hValAlaGluGlnAsnLysAlaAspPheThrSerLysAspPheLysGln 381
1491 CAGCTTTAGAC.....GCTGCTCAATTAACAACAGATTCTAAAGGTGA 1534
382 AsnAsnLysValTyrGluGlyAspAsnAlaTyrValGlnValGluLeu 398
1535 GGTACATTGACAGTTTCTGTACTAATGACGTGAAGCCAGTGTATA 1584
398 sAsp.....GlnPheAsnAla 403
1585 TGAATCTACACACACTAACATAGTACTTCAAATAAAATATATGTCAT 1634
404 .....ValThrThrGlyLysValGluTyrGluSerLeuAsnThr 416
1635 CTGCTTTACAAACTACTGCTTCTAAAGTAAGTTCGCTGCTCTCAAGCA 1684
417 GluValAlaVal.....ValAspLysAlaThrGlyLysValThr 429
1685 GAGTATACATATGAGTAACTGCTGCTGATATGCT...GGAGAGTTC 1731
429 rValLeuSerAlaGlyLysAla..ProValLysValThrValLysAsp 444
1732 TGCATTTGCGCTACTAAGCGTCCGGAATACAAAGTTATTTGTAAGA 1781
445 .....SerLysGlyLysAlaLeuValSerHisThrValGln 456

```

```

1782 AAGCTTTAACTTACTTAAATAATGAAATCGTTAATGTTGATTCATTA 1831
456 uIleGlu..... 458
1832 GATTAAGATCGTGTATTTCAACACTTACAAATGCTAAATTCGTTGAT 1881
459 .....AlaPheAlaGlnLysAlaMetLysAspIleLys..L 470
1882 TGATCCAGATACTGCAGTATACTTACAGCGCATTAACCAAAACAAAT 1931
470 euGluLysThrAsn..... 474
1932 CTGTAAACAAATGATTAAGGTGAAGCTTACATTTGTTATCGTCTGAT 1981
475 .....ValAlaLeuSerThrLysAspValThrAspLe 485
1982 ACAGTAACGATTAATGCAACACCAATTCCTTGATGATATTAATCTTC 2031
485 uLysValLysAlaProValLeuAspGlnTyrGlyLysGluPheThrAla 502
2032 TGATCAAAACCAAGCGACCTGATGAAGGTGAACCAAAAGCATTCGAC 2081
502 rovalThr...ValLysValLeuAspLysAspGlyLysGluLeuLysGlu 517
2082 CAATCTTACTTCCAGCACCATATCTTGATGGCTCAGCTGACGTAAAGCA 2131
518 GlnLysLeuGluValAlaLysTyrValAlaAsnArgLysLeuValLeu 534
2132 TACAAA.....AAATCAGATCTTAATTAAGCTGTAACTAAGTTGA 2172
534 aGlyGlnGluAlaGlyAsnTyrThrValValLeuThrAlaLysSerGlyG 551
2173 TGGTCTCAAACTGCAGATTTTGCAGCAGAAATTAACCAAAAGCGGCA 2222
551 LuLysGluAlaLysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAla 567
2223 AAAAAGTACTGCTTCTTAT.....AAGAAAGCAACT 2257
568 PheSerLysPheGluValArgLysAspThrGluLeuAspLysTyrVal 584
2258 TATACATCTCAATATGCTGCTTAATGATTAATTAAGTAGAATACCAAG 2307
584 lThrGluLysnGlnLysAsnAlaMetThrValSerValLeuProValA 601
2308 TATCTCACCAATCTGACTACACAGTAACTTAGCAACT...ACTTAT 2354
601 sPalaAsnGlyLeuValLeuLysGlyAlaGluAlaGlu..... 614
2355 CTTCACAGGAAGTGTATTACACCTGCTTAAGAATTTGAAGTTACTTCA 2404
615 .....LeuLysValThrThr 620
2405 GTGATGGTAAACAACTGCTGTTAAAGTAATGTACTACAGATTTGCTG 2454
621 AsnLysGluGlyLysGluValAspAlaThrAspAlaGlnValThrVal 636
2455 TAAATACAGACGTAAAGACTATGCAATTACTGTAAAGAACCTACAGCTA 2504
636 ..... 636
2505 CATTACAGCTACAAATGAACTTCCAAACTCTTACACTGGGTGACTACT 2554
637 .....GlnAsnAsnSerValIleThrVal 644
2555 CAATTCAATACAGCTGATTTGTTCAACACGCAACTATTTGGTTTGC 2604
644 lGlyGln.....GlyAlaLysAlaGlyGluThrTyrLys 655
2605 TGTATAAAACCCAGTGAATATGCTGTATCAGGCACAAACATATTAAT 2654
655 ..... 655
2655 ACTTCGAGCTAATGTAATGAACTATTGGTGAAGCGGATGGGAAGCA 2704

```



```
1346 AAGAACGTTACTGTAACCTGTAACCAATATTAACCTGTAATCAATATGTT 1395
160 LysThrValThr.....LysAlaGluAlaAlaGlnPheIleAlaLysTh 174
|||||.....:|||||
1396 AAACACTTAAGTGGTTACAAATGACAGAGCTTACGAGTTAGTG..... 1437
174 rAspLysGlnPheGlyThrGluAlaAlaLysValGlnSerAlaLysAlaV 191
|||.....:|||||
1438 .....TTCACTGCACAAACGATCAACACCAACTGTTGCTACCGGCTC 1477
191 alThrThr..... 193
|||||
1478 CTACTACTTTAGGTGGTACAACTTATCTAGTCTTCTTCAACAAT 1527
193 ..... 193
1528 GTTTGGGGAATTAATGGCTGGTGTGAATGAAGCTGAACCTTATTATCC 1577
194 ...GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
|||.....:|||||
1578 TGGCTCTTCAATTCACMACACCTTGGCTACTAAGTTAGACGAATCTACTT 1627
209 ySGluAspIleLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
|||.....:|||||
1628 TAGCTGATTAAGTCTTGTATTAGTTGAAGAAATCTGGTACAGTGTGCT 1677
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
|||.....:|||||
1678 TCTGAACCTAAATTAATGACAGCCCTGAATGTAACCTTATAGTCCAAA 1727
242 rSerAsnLeuAlaAlaLysGlnThrThrValAspValAsnLysValG 259
|||||.....:|||||
1728 ACCGGACCTTAAGAAATATCAATCTCAATCAAAATTAATAAAGGCT 1777
259 LysThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
|||||.....:|||||
1778 TGAAGTCCCATTAAGGATTAATGATAGCAGCTTAACGAGAAACATTT 1827
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnP 290
|||||.....:|||||
1828 GAGTTCAAACCTCAAGACTTAACCTGCT.....CCTACAGTTATTAAGCT 1871
290 eThrValLysAsp...GluAsnGlyThrGluValValSerProGluGlyI 306
|||||.....:|||||
1872 AACGCTTAATAAATGGCGAGCTGGATTAAGTAAGTAAGT.....GAAGCTC 1915
306 LeuGlnPheValThrProAlaAlaGluLysIleAsnAlaLysGluIle 322
|||||.....:|||||
1916 AAGAAATTAAGTGAAGTCTCAGAGAAATTTAAATTAATTAATGCTACA 1965
323 ThrLeuAlaLysGlyThrSerThr..... 330
|||||.....:|||||
1966 ACCGTTTCGGTAGCACATCAATACGTCGAAGTTGCTGTAGTAAGGC 2015
331 .....ThrV 332
2016 GGGTCAAACTTATCTGCTTACAGCAAGTGACATCATTTCCAGCTAGG 2065
332 alLysAlaValTyLysLysAspGly.....LysValValAla 344
|||||.....:|||||
2066 TTGAAGCCGTTACTGCTCAAGATGGAACATCAAAAGTGAAAGTCTCTCT 2115
345 GluSerLysGlu..... 348
|||||
2116 AACCAATTAGAACGTAAACCAAGGTACAAATTAAGTTCGTTGCTAAAGG 2165
349 .....ValLysValSerAlaGluGluValAlaValAlaSerI 361
|||||.....:|||||
2166 TGCACACACTCTCTTTAAAGATGCTGCAAAATGCAAAATCTTACGACACTA 2215
361 LeSerAsnTyrThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
|||||.....:|||||
2216 ACTATATCTATCTTACAACTGAAGGTCAA...GACGTAAACGACCA 2262
378 AspPheLysGlnAsnAsnLysValTyGluGluLysAsnAlaTyValGI 394
|||.....:|||||
2263 ACGGTT.....ACAAAGTATTCAAAGCTGATTTCT..... 2292
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
|||||.....:|||||
2293 .....TTAAAGACGCT...GATGCAGTTACTACA..... 2319
411 yRGluSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
2319 ..... 2319
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
|||.....:|||||
2320 CTTACGAACGTTGATGACAGCTCAA.....AATTCATCATCCAAATT 2360
443 .....LysAspSerLysGlyLysAlaLeuValSerHisTyrV 455
|||.....:|||||
2361 TAGCGAAGAAATTAATAAAGTCTAGTGT...TCTTATGCTGGGTGGCAAG 2407
455 alGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
|||.....:|||||
2408 TTAAGTGTGAGAAATTAACAAACGATGGGTAGAT.....GCT 2448
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLy 488
|||||.....:|||||
2449 GGTACTGGAACAACTGATCAGTT..... 2472
488 sAlaProValLeuAspGlnTyGlyLysGluPheThrAlaProValThrV 505
|||||.....:|||||
2473 .GCTCTTAAGACAGATGCCAAATGTAAGTAAACACTGCTGTGTACTAT 2521
505 alLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
|||.....:|||||
2522 TAACGGGCTTGGACATAACGACAAAGATGGGAA...TTGCGTCTGTA 2568
522 AlalysTyValAsnArgGluLeuValLeuAsnAlaAlaGlyGlu..... 536
|||.....:|||||
2569 GTAGTAAGTGTCTTACGATGATGATTCGTGATGACGTGTAATGTAAAT 2618
537 .....GluAlaGlyAsnTyThrValLy 544
2619 TAGGAAAAAGATATTTAATTCGTTACACAGCTGGAGACACCTGTAG 2668
544 alLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
|||.....:|||||
2669 CTTCGTGAAAGCTGCTGCTGACAAAGATGGTCAA..... 2703
561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
|||.....:|||||
2704 .....AACGCTTCGCTGATTCCTCA.....ACAAG 2729
577 pThrGluLeuAsp.....LysTyValThrGluGluAsnGlnLys 591
|||.....:|||||
2730 CACTGCAATTAATACAACTAAGAGCTTATTAAGTTGAATTCATGA... 2775
591 snAlaMetThrValSerValLeuProValAlaAspAlaAsnGlyLeuValLeu 607
|||.....:|||||
2776 .....ACTGATTTAGCGGAAGTTAAACCTGAAACATCGTGT 2814
608 LysGlyAlaGluAlaAlaGluLeuLysValThrThrAsnLysGluGI 624
|||.....:|||||
2815 AAAGATGACAGCAGTAATGTCGAGTGTCTGCTAACACCATTAAGACGG 2864
624 yLysGluValAspAlaThrAspAlaGlnValThrValGlnAsnAsnSerV 641
|||.....:|||||
2865 TTCT.....ACAAATAAATTTG 2881
641 alIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTyLysValThr 657
|||||.....:|||||
2882 TATTCACCTCCATCTCAAGAAATTAAGCTGGTACAGATTACTCTGTACCA 2931
```

```

658 ValValLeuAspGly.....LysLe 664
    : : : : :
2932 AT.....GACGGTGAGAGATAAGTAGTAACAACTCTAAATA 2975
664 uileThrThrHisSerPheLysValValAspThrAlaProThrAlaLysG 681
    ||||| ||||| ||||| ||||| |||||
2976 CATTAAT.....TCGTTCAAGACTGATTCGCGAATCCAAAG..... 3012
681 LysLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProAsn 697
    : : : : : ||||| ||||| |||||
3013 .....TTATCTCAATCAACGACTTGTGACGGTGCGAGTTTAC 3048
698 AlaAspLeuLysAlaAlaLeuLeuAsnLeuSerValAspGlyValPr 714
    : : : : : ||||| ||||| |||||
3049 GTTGAC...CGTCTCAAAACAATTGAAATTCACAGCTTCAGTTCC 3095
714 o...AlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSer 730
    ||| : : : : : ||||| ||||| |||||
3096 AAACCCAAACATCACTCTTAAGAAAGCTGACGGAACCTTCATTACTAAT 3145
730 LaasP...ThrAsnValValAlaGluAsnGlyThr.....ValGly 742
    : : : : : ||||| ||||| |||||
3146 ACACCTTAGTAATAGTAATATGAAATTAACATCAAAATGTATTC 3195
743 AlaLysGlyAlaThr.....SerLetyrValLysAsnLeuThrValVa 757
    ||||| : : : : : ||||| ||||| |||||
3196 CACAAGGCTGTAACACTGACGAGTTTACTCAATATGAGTTCAGCTTC 3245
757 LysAspGlyLys.....GluGluLysValGluPheAspL 769
    ||||| : : : : : ||||| ||||| |||||
3246 AAAAGATTTTCAAACTGCTACTGATATGATGACGAAGCTTCATCTC... 3291
769 ysaLysAlaGluValAlaValSerLetyrLysGluAlaLysProAla 783
    : : : : : ||||| ||||| ||||| |||||
3292 .....ATCACAGGTTCTGTGCTACTGACGAAGTAAACCTGCT 3330
seq_name: /SIDS2/gcgdata/geneseq/NA1999.DAT:AAx22748
seq_documentation_block:
ID AAX22748 standard; DNA; 3687 BP.
XX AC AAX22748;
XX DT 24-AUG-1999 (first entry)
XX DE B stearothermophilus sbsA DNA.
XX KW sbsA protein: S-layer protein; Gram-negative; prokaryotic host cell;
    integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
    eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
    immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
    molecular laser; universal carrier molecule; monomolecular layer; ds.
XX OS Bacillus stearothermophilus.
XX FH Key Location/Qualifiers
    CDS 1..3687
    FT /*tag= a
    FT /product= "sbsA"
    FT sig_peptide 1..90
    FT /*tag= b
    FT mat_peptide 91..3684
    FT /*tag= c
XX DE19732829-A1.
XX PD 04-FEB-1999.
XX PF 30-JUL-1997; 97DE-1032829.
XX PR 30-JUL-1997; 97DE-1032829.
XX PA (LUBITZ) LUBITZ W.

```

```

XX LUBITZ W, Resch S;
PI WPI; 1999-122189/11.
XX P-PSDB; AAW93252.
DR
XX
PT Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
PT integrated into membranes or organelles or secreted into periplasma
PT or growth medium, and nucleic acid encoding S-layer proteins with
PT peptide insertions, used in vaccines or for enzymatic reactions
XX
PS Claim 5; Page 9-14; 34pp; German.
XX
CC This invention describes a method for the production of a S-layer protein
CC (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
CC transformed with nucleic acid (II) encoding (I), linked to a signal
CC sequence (SS) that encodes a protein which causes at least one of (i)
CC integration of (I) into the external or cytoplasmic membranes and/or
CC (ii) secretion of (I) into the periplasmic space or extracellular medium,
CC (b) culturing the cell to express (I) and (c) optionally recovering (I)
CC from the membranes, periplasmic space and/or extracellular medium.
CC Alternatively, a eukaryotic cell is used as host and then the SS, which
CC is optional, promotes integration into the cytoplasmic membrane or an
CC organelle and/or secretion into the extracellular medium. (I), and
CC derived structures, may include a wide variety of polypeptide inserts and
CC are useful as (i) vaccines or adjuvants (with immunogenic epitopes or
CC immunostimulants) inserts such as cytokines (ii) as receptors (inserts
CC are enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a
CC 'molecular spinnerette' for production of PHB or luciferase for use as
CC molecular laser (when combined with substrate and oxygen) and (iii) as
CC universal carrier molecule (streptavidin is inserted) for use in
CC hybridisation and immuno assays, or for selective elimination of
CC cytokines, toxins etc. from body fluids (inserts are specific binding
CC epitopes). In this system, heterologous (I) do not form inclusion bodies
CC but rather monomolecular layers, and in eukaryotic cells they undergo
CC glycosylation. This sequence encodes the Bacillus stearothermophilus
CC sbsA protein which is used to illustrate the method of the invention.
XX
SQ Sequence 3687 BP; 1316 A; 660 C; 715 G; 996 T; 0 other:
XX
alignment_scores:
Quality: 248.50 Length: 898
Ratio: 0.597 Gaps: 44
Percent Similarity: 46.325 Percent Identity: 22.383
alignment_block:
US-09-754-947-1 x AAX22748 ..
Align seg 1/1 to: AAX22748 from: 1 to: 3687
55 AsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGlnG 71
    : : : : : ||||| : : : : :
967 GATGTTCTTATCTACAGACGAAACATCACTGATGCTCA.. 1014
71 yGlnTrpTrThrProPheIleAlaValGluLysAlaGlyValIlel 86
    ||||| : : : : : ||||| ||||| |||||
1015 .....ACTCCATTCGAAATAATATACGAGTATTAAGTAGTCTTA 1054
88 ysgLYThrGlyAsnGlyPheGluProAsnGlyLysIleAspRValSer 104
    ||||| : : : : : ||||| ||||| |||||
1055 AAGGTATTAA.....GACAAAAATGCGCAAGAATTAAAGACAGAT 1095
105 MetAlaSerLeuLeuValGluAlaTyrlLysLeuAspThrLysValAsnG 121
    : : : : : ||||| : : : : : ||||| |||||
1096 GCATTCACCTTCAAGCTTCGAAATGATGCTGACTAGTCAAGTGTGG 1145
121 yThrProAlaThr..... 125
    ||||| : : : : : ||||| ||||| |||||
1146 AACTAATGTACAAACAACTCTGTAACTAGCAGCAGCTACTTTCG 1195
126 .....LysPheLysAspLeuGlu 131
    ||||| : : : : : ||||| ||||| |||||

```

```
1196 ACACCTAGCAGTACTTAAACAGTAGTATTGATAGTTGTTAGCACCTGAA 1245
132 ThrLeuAsn.....TrpGlyLys.. 137
|||||
1246 ACTGTAAACGCTCGAAGCTTACTATTACAGATGTTGAAACGGAAG 1295
138 .....GluLysAlaAsnIleLeuValGluLeug 147
|||||
1296 CATTCCAGTAATGTCATCTACTTGTCTGTTCTACAAATTACTATTACGTTAA 1345
147 LylleSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159
|||||
1346 AAGAACGCTTAGTAAGTACTGCTAAACATATAAACTCTCTATCAATATGTT 1395
160 LysThrValThr.....LysAlaGluAlaIleGlnPheIleAlaLysThr 174
|||||
1396 AAACATTAATACGTGTTACAAATGAGAGCTTACGAGTAGTG..... 1437
174 rAspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlav 191
|||||
1438 .....TTCACGCAAGCAGCACACCACCACTGTTGCTACGCGTC 1477
191 aLThrThr..... 193
|||||
1478 CTACTACTTTAGTGTGACAACTTATCTACTGTTCTCTTACACCAAT 1527
193 ..... 193
1528 GTTTGGGCTAAATTGGCTGTGTGATGAAGTGAAGCTGAACCTTATTATCC 1577
194 ....GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
|||||
1578 TGGCTCTTAATTCACAAACAGCTTGTCTACTAAGTTAGACGAATCTACTT 1627
209 yGluGlnLleLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
|||||
1628 TAGCTCATTAAGTTGTTGATTAGTTGAAAAAAGATCTGGTACAGTTGTCGT 1677
226 LysGluValThrLeuSerGlnAspLysArgSerAlaThrValGluLeuTy 242
|||||
1678 TCTGAAGCTAAATATATATGACAGCCTAAATGTGTAACTTATAGTGCCAA 1727
242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
|||||
1728 ACCGCACTTAAAGAAATACATCTATCAATCAAAATTAAGAAAGGCT 1777
259 lYlSerThrGlu.....ValAlaValIleSerLeuGlnAlaLysThrIle 273
|||||
1778 TGAAGTCCCATTAAGGTATTGAATTAGGACCTGTTACGAGAAACATAT 1827
274 GluMetAlaAspGlnThrValAlaAlaAspGlnProThrAlaLeuGlnPh 290
|||||
1828 GAGTTCAAAACATCAAGACTTAAGTCT.....CCTACAGTTATTAGCGT 1871
290 eThrValLysAsp...GluAsnGlyThrGluValAlaSerProGluGlyI 306
|||||
1872 AACGCTTAAAAATGGCGAGCTGGATTAAGTAAGTACT.....GAAGCTC 1915
306 lGluPheValThrProAlaIleGluLysIleAsnAlaLysGlyLeuIle 322
|||||
1916 AAGATTTACTCTGAAGTTCTCAGAGAAATTTAAATATACATTTAATGCTACA 1965
323 ThrLeuAlaLysGlyThrSerThr..... 330
|||||
1966 ACCGTTTCGGTAGCACATACATACGCTCAAGTGTGCTAGTAAAGC 2015
331 .....ThrV 332
|||||
2016 GGGTCAAACTTATCTGCTTACAGCAAGTGACATCATTCACAGCTAGTG 2065
332 aLlValAlaValTyLysLysAspGly.....LysValAlaAla 344
|||||
2066 TTGAAGCGTTTACTGCTCAAGATGGAACATACAAAGTGAAAGTCTGCT 2115
345 GluSerLysGlu..... 348
|||||
2116 AACCATATTGAACGTAAACCAAGGTAAACAATTAAGTAGTGTGGTAAAG 2165
349 .....ValLysValSerAlaGluGlyAlaAlaValAlaSerI 361
|||||
2166 TGCACACGCTCCTGTTAAAGATGCTGCACAAATGCAATCTTTAGCACTA 2215
361 lSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
|||||
2216 ACTATATCTATTCATTTACAACTGAAGGTCAA..GACGTAAACACACACA 2262
378 AspPheLysGlnAsnAsnLysValTyrgLugLysAspAsnAlaTyValGl 394
|||||
2263 ACGGTT.....ACAAAGTATTCAAAGGTGATTC..... 2292
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
|||||
2293 .....TTAAAGACGCT..GATGCAGTTACTACA..... 2319
411 yrgLuserLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
2319 ..... 2319
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
|||||
2320 CTTACGAACGTTGATGCAGGTCAA.....AAATTCATCATCCAAAT 2360
443 .....LysAspSerLysGlyLysAlaLeuValSerThrIsthrV 455
|||||
2361 TAGCGAAGAAATTAAAACTTCTAAGTGT...TCTTACTGTGGTGCAAG 2407
455 aIcGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
|||||
2408 TTAAGTGTGAGAAATTACAAACACGAGTGGTAT.....GCT 2448
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLy 488
|||||
2449 GGTACTGGAACAACTGTATCACTT..... 2472
488 sAlaProValLeuAspGlnTyrgLysGluPheThrAlaProValThrV 505
|||||
2473 .GCTCTTAAGACAGATGCAAATGTGAAGTAAACACCTGCTGTGTATAT 2521
505 aLlValValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
|||||
2522 TAACTGGTCTTGACAAATTAACGACAAAGATGCGAAA..TTGCGTGTGTA 2568
522 AlalTyTyValAsnArgGluLeuValLeuAsnAlaIleGlyGln..... 536
|||||
2569 GTAGTAAAGTCTTACTGATGAGATTTGCTGATGAGCTGGAATGTAAT 2618
537 .....GluAlaGlyAsnTyThrValLy 544
|||||
2619 TAAGAAAAAGATATTTAATTCGTTACACAGCTGGAGACACACTGTAG 2668
544 aLleuThrAlaLysSerGlyGlnLysGluAlaLysAlaThrLeuAlaLeu 560
|||||
2669 CTTCTGTAAAGCTGCTGCTGACAAAGATGCTCAA..... 2703
561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
|||||
2704 .....AACGCTTCGCTGATTCOCA.....ACAG 2729
577 PThrGluLeuAsp.....LysTyTyValThrGluGlnAsnGlnLysA 591
|||||
2730 CACTGCATTTGATACAACTAAGAGCTTATAGTTGAATTCATGA... 2775
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607
|||||
2776 .....ACTGATTTAGCGGAAGTTAAACCTGAGAACATGCTGTT 2814
```

```
608 LysGlyAlaGluAlaGluLeuLysValThrThrAsnLysGluG1 624
    ||| |||
    : : :
    : : :
2815 AAAGATGACGAGCATATGCGTACTGCTGACGATTAAGACGG 2864
624 YLysGluValAspAlaThrAspAlaGluValThrValGlnAsnAsnSerV 641
    |||
    : : :
2865 TTCT.....ACAAATAAATTG 2881
641 AlIleThrValGlyGlnGlyAlaLysAlaGlyLutThrTyrLysValThr 657
    |||
    : : :
2882 TATTCACCTCAGATTCACAGAAATTAAGCTGGTACAGTTACTGCTGAACA 2931
658 ValValLeuAspGly.....LysLe 664
    : : :
2932 ATT.....GACGGTGTGAGAGATTAAGTAGTAAACAAATCTCTAAATA 2975
664 ULleThrThrHisSerPheLysValValAspThrAlaProThrAlaLysG 681
    |||
    : : :
2976 CATTAAT.....TCGTTCAAGACTGTATCTGCGAATCCAAACG..... 3012
681 LylLeuAlaValGluPheThrSerThrSerLeuLysGluValAlaProAsn 697
    |||
    : : :
3013 .....TTATCTTCATACAGCAATTCGTCGACGGTGCAGTTAAC 3048
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValPr 714
    |||
    : : :
3049 GTTTCAC.....CGTTCTAAACAAATTAACATTAATGCAATTCAGCTGCTCC 3095
714 O...AlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerA 730
    |||
    : : :
3096 AAACCCAACTACCTCTTAAGAACGCTGACGAGCTTCATTTACTAAT 3145
730 laasp...ThrAsnValAlaAlaGluAsnGlyThr.....ValGly 742
    |||
    : : :
3146 ACACCTTAGTAATCTAATATGAATTAACATTAACAAATGTAATTC 3195
743 AlaLysGlyAlaThr.....SerIleTyrValLysAsnLeuThrValVa 757
    |||
    : : :
3196 CACAAAGGTGTACACTGACGAGTTTACTCAATATGAGTTAGCGATTTC 3245
757 LlysAspGlyLys.....GluGlnLysValGluPheAspL 769
    |||
    : : :
3246 AAAGATTTTCAAACTGCTACTGATGATTTGATGACAAAGTTACTATTC... 3291
769 ysaAlaValGlnValAlaValSerIleLysGluAlaLysProAla 783
    |||
    : : :
3292 ....ATCACAGGTCTGCTGCTACTGACGAGAAATGAACCTGCT 3330
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA71797
seq_documentation_block:
ID AAA71797 standard: DNA: 3687 BP.
XX
AC AAA71797:
XX
DT 08-JAN-2001 (first entry)
XX
DE B. steaerothermophilus sbSA DNA.
XX
KM S-layer gene: sbSA; antibacterial; vaccine; adjuvant; bioreactor;
XX polyhydroxyalkanoate synthesis; ds.
XX
OS Bacillus steaerothermophilus.
XX
FH Key Location/Qualifiers
FT CDS 1..3687
FT FT /*tag= a
FT FT /product= "sbSA"
FT FT sig_peptide 1..93
FT FT /*tag= b
FT FT mat_peptide 94..3684
FT FT /*tag= c
XX
```

```
PN DE19903345-A1.
XX
XX 03-AUG-2000.
XX
PF 28-JAN-1999; 99DE-1003345.
XX
PR 28-JAN-1999; 99DE-1003345.
XX
XX (LUBI/) LUBITZ W.
XX
XX Lubitz W;
XX
XX WPI; 2000-533868/49.
XX
XX P-P-SDB; AAB10625.
XX
XX Host cell, useful e.g. as bioreactor for production of
XX poly(hydroxyalkanoate), containing two or more recombinant
XX polypeptides, with at least one in carrier-bound form -
XX
XX Disclosure; Page 8-12; 26pp; German.
XX
XX This invention describes a novel host cell (A) comprising at least two
XX functional recombinant polypeptides (I), at least one being in carrier
XX bound form. The products of the invention have antibacterial activity.
XX (A), or, where bacterial, their ghosts (B), are useful as vaccines or
XX adjuvants (specifically for presentation of immunogenic epitopes of
XX pathogens or autoantigenic immunostimulatory polypeptides, e.g. cytokines),
XX or preferably, as enzyme reactors for performing a cascade of reactions,
XX specifically synthesis of poly(hydroxyalkanoate). Localization of
XX individual (I), specifically enzymes, in separate cellular compartments
XX avoids adverse reactions between products and substrates, when being used
XX as bioreactors. (I) can be produced in carrier-bound form without loss of
XX function. This sequence encodes the Bacillus steaerothermophilus S-layer
XX protein sbSA which is used to illustrate the method of the invention.
XX
XX Sequence 3687 BP; 1316 A; 660 C; 715 G; 996 T; 0 other:
XX
XX alignment_scores:
XX Quality: 248.50 Length: 898
XX Ratio: 0.597 Gaps: 44
XX Percent Similarity: 46.325 Percent Identity: 22.383
XX
XX alignment_block:
XX US-09-754-947-1 x AAA71797 ..
XX
XX Align seg 1/1 to: AAA71797 from: 1 to: 3687
XX
XX 55 AsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGlnG1 71
XX : : : : : ||| : : : : :
XX 967 GATGTTCTTATATCTACAGAGGAAACCTATCACTGAGTGCCTCA.. 1014
XX
XX 71 YGlnTrpTyrThrProPheIleAlaValGluLysAlaGlyValIleL 86
XX : : : : : ||| : : : : :
XX 1015 .....ACTCCATTCCGAATAATATACGAGATTAAGTAGTAGTGA 1054
XX
XX 88 ysgLyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArgValSer 104
XX ||| : : : : : ||| : : : : :
XX 1055 AAGGTATTAA.....GACAAAAATGCAAAAGATTTAAAGAACAGAT 1095
XX
XX 105 MetAlaSerLeuLeuValGluAlaTyrIleLysLeuAspThrLysValAsnG1 121
XX : : : : : : : : : : :
XX 1096 GCATTCACCTTCAGCTTCGAATGATGCTGATTAAGTTACTCAAGTGTGG 1145
XX
XX 121 YThrProAlaThr..... 125
XX ||| : : : : : ||| : : : : :
XX 1146 AACTATATTAACAACAACAACTTCTGTAACCTTAGCAGCAGACTTTGCG 1195
XX
XX 126 .....LysPheLysAspLeuGlu 131
XX ||| : : : : : ||| : : : : :
XX 1196 ACACGTACGATACCTTTAAGACGATGATTTGATTAAGTTGTTGACACCTGA 1245
XX
XX 132 ThrLeuAsn.....TrrpGlyLys.. 137
```

```

1246 ACTGTAAACGCTCGAACGTTACTATTACAGATGTTGAAACTGGAACG 1295
138 .....GluysAlaasnIleLeuValGluLeu 147
1296 CATTCCAGTAATTGCATCTACTTCTGTTCTACAAATTACTATTACGTTAA 1345
147 LylleSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159
1346 AAGAACGCTTAGTAAGTGGTAACCAATATATACTGCTACTCATATATGTT 1395
160 LysThrValThr.....LysAlaGluAlaAlaGlnPheIleAlaLysThr 174
1396 AAAACATTAACCTGTTACATGCGAAGCTTACGAGTGTG..... 1437
174 AspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaVal 191
1438 .....TTCACGCAAAACGCAATCAGCACCACTGTTGCTACCGCTC 1477
191 aIThrThr..... 193
1478 CTACTACTTTAGGTGGTACAACTTATCTACTGTTCTCTACACAAT 1527
193 ..... 193
1528 GTTGGGGAATATTGGCTGGTGGTGAATGAAGCTGAACCTTATTATCC 1577
194 .....GlnLysValGluValLysPheSerLysAlaValGluLysLeuThr 209
1578 TGCTCTTCAATTACCAACAACGTTTCTCTACTAAGTAGAGAACTTACTT 1627
209 yAsGluAspIleLysValThrAsnLysAlaAsnAspLysValLeuVal 225
1628 TAGCTGATAACTTGTATAGTAGTGAAGAAATCTGCTACAGTGTGTGCT 1677
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
1678 TCTGAACCTAAATATATATAGCAGCGCTAAATGTACTTATGTGCGCAA 1727
242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
1728 AGCGACCTTAAAGAAATATACATCTATCAATCAAAATTAAGGCT 1777
259 LysThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
1778 TGAAGTCCCATTAAGCTATTGAATTAGCGACTGTAAACGAGAAACATAT 1827
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnPh 290
1828 GAGTTCAAACTCAAGACTTAACGTCT.....CCTACAGTTATTAGCGT 1871
290 eThrValLysAsp.....GluAsnGlyThrGluValValSerProGluGlyI 306
1872 AAGCTTAATAAATGGCGACGCTGATTAAGTAACT.....GAAGCTC 1915
306 leGluPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIle 322
1916 AAGAAATTACTGTGAAGTCTCAGAGAAATTAATACATTATTAAGCTACA 1965
323 ThrLeuAlaLysGlyThrSerThr..... 330
1966 ACCGTTTCGGTAGCACACATACATACGTCAGTTGCTGTAGTAAACG 2015
331 .....ThrV 332
2016 GGTGCAAACTTATCTGCTTACAGCAAGTGACATCATTCAGCAGTAGG 2065
332 aLysAlaValTyLysLysAspGly.....LysValValAla 344
2066 TTGAACCGCTTACTGTCAGAGATGCAACATCAAGTGAAGAGTTCCTGCT 2115
345 GluSerLysGlu..... 348

```

```

2116 AACCAATTAGAACGTAAACCAAGGTACAAATTAGTAGTGTGCTAAAGG 2165
349 .....ValLysValSerAlaGluGlyAlaAlaValAlaSerI 361
2166 TGCACAGCTCCTGTTAAAGATGCTGCAATGCAATGCAATCTTTAGCACTA 2215
361 leSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
2216 ACTATATCTATACATTTACACTGAAGCTCAA...GAGCTAACACACACA 2262
378 AspPheLysGlnAsnAsnLysValTyGluGlyAspAsnAlaTyValG 394
2263 ACGGTT.....ACAAAGTATTCAAAGCTGATTC..... 2292
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
2293 .....TTAAAGACGCT...GATGAGTTACTACA..... 2319
411 yrcLusSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
2319 ..... 2319
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal.... 442
2320 CTTACGAACGTTGATGACAGTCAA.....AATTCACATCAAT 2360
443 .....LysAspSerLysGlyLysAlaLeuValSerHisThrV 455
2361 TACGAGACATTTAAAACTCTTAGTGT...TCTTAGGCGGTGCGCAAG 2407
455 aLcLulIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
2408 TACAGTGCAGAAATTACAAACACGAGTGGTGAT.....GCT 2448
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysVally 488
2449 GGTACTGGAACAACGTATCACTT..... 2472
488 sAlaProValLeuAspGlnTyGlyLysGluPheThrAlaProValThrV 505
2473 .GCTCTAAGACAGATGCAATGTTAAAGTAACTGCTGTGTTACT 2521
505 aLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
2522 TAACTGCGCTTGACAATAACGACAAAGATGCGAA...TTGCGTCTGTA 2568
522 aLalysTyrValAsnArgGluLeuValLeuAsnAlaAlaGlyGln.... 536
2569 GTAGATTAAGCTTCTACTGATGGAATTCGTGATGCTGTAATGTAT 2618
537 .....GluAlaGlyAsnTyThrValV 544
2619 TAAAGAAAAGATATTAAATTGTTTACACAGCGTCGAGACACACTGTAG 2668
544 aLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
2669 CTTCTGTGAAGCTGCTGCTGCAAAAGATGTCAA..... 2703
561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
2704 .....AACGCTTCTGCTCATTCCTCA.....ACAAAG 2729
577 pThrGluLeuAsp.....LysTyrValThrGluGluAsnGlnLysA 591
2730 CACTGCATTTGATACACTAAGAGCTTATTAGTTGAATTCATGAA.... 2775
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607
2776 .....ACTGATTACCGGAAGTTAAACCGCAAGAACATCGTTGTT 2814
608 LysGlyAlaGluAlaAlaGluLeuLysValThrThrThrAsnLysGluG 624
2815 AAGATGCAAGCAAGTAATGCGTAGCTGCTGCTACTGTAAACGATTAAGCGG 2864

```

```

624 yLysGIuValAspAlaThrAspAlaGlnValThrValGlnAsnSnserv 641
    :::::  ||:::
2865 TTCT.....ACAAATAAATTG 2881
641 allIeThrValGlnGlnValAlaLysAlaGlnThrTyrLysValThr 657
    :::::  ||:::
2882 TATTCACTCCATCTCAAGATTAAAGCTGGTACGTTACTCTGTACA 2931
658 ValValLeuAspGly.....LysLe 664
    :::::  ||:::
2932 ATT.....GACGGTGCAGAGATAAGTAGTACACACTCTCAATA 2975
664 uIleThrThrHisSerPheLysValAlaAspThrAlaProThrAlaLysG 681
    |||||  |||||
2976 CATTAAT.....TCGTCAGAGACTGTATCTGCAATCAACG..... 3012
681 LysLeuAlaValGluPheThrSerThrSerLeuLysGlnValAlaProAsn 697
    :::::  ||:::
3013 .....TTATCTTCATACACATTCGCTGACGGTGCAGTTTAC 3048
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValPr 714
    :::::  |||||
3049 GTTGAC...CGTTCTAAACAATATACATTCGATTCAGCATTCGTTCC 3095
714 o...AlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerA 730
    |||  :::::
3096 AAACCAACAATCACTCTTAAGAACGCTGACGGAACCTTACTTAAT 3145
730 laAsp...ThrAsnValAlaGlnAsnGlyThr.....ValGly 742
    :::::  |||||
3146 ACACCTTGTAATGTAAATGTAAATGAAATTAACCTACAAATTTGTTATTC 3195
743 AlaLysGlnAlaThr.....SerIleTyrValLysAsnLeuThrValVa 757
    |||||  :::::
3196 CACAAGGTGTAACACTTGACGAGTTTACTCAATGATGATTAGCAGTTTC 3245
757 lLysAspLys.....GluGlnLysValGluPheAspL 769
    |||||  :::::
3246 AAAAGATTTCCAACTGCTACTGATATTGATAGCAAAAGTTACATTTC... 3291
769 ysaIaValGlnValAlaValSerIleLysGlnAlaLysProAla 783
    :::::  |||||
3292 ....ATCACAGGTTCTGTTGCTACTGACGAGATAAACCTGCT 3330
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAx78246
seq_documentation_block:
ID AAX78246 standard; DNA: 3768 BP.
AC AAX78246;
XX
XX
XX 24-AUG-1999 (first entry)
XX
DE B. stearothermophilus sbas/bacteriophage fd Gene 3 fusion protein DNA.
XX
XX sbas protein: S-layer protein; Gram-negative; prokaryotic host cell;
XX integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
XX eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
XX immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
XX molecular laser; universal carrier molecule; monomolecular layer; Gene 3;
XX fusion gene; ds.
XX
XX synthetic.
XX OS Bacteriophage fd.
XX OS Bacillus stearothermophilus
XX DE19732829-A1.
XX
XX 04-FEB-1999.
XX
XX 30-JUL-1997; 97DE-1032829.
XX

```

```

PR 30-JUL-1997; 97DE-1032829.
XX
XX (LUBITZ/) LUBITZ W.
XX
XX Lubitz W, Resch S;
XX
XX WPI; 1999-122189/11.
XX
XX Producing S-layer proteins in Gram-negative bacteria of eukaryotes
XX integrated into membranes or organelles or secreted into periplasma
XX or growth medium, and nucleic acid encoding S-layer proteins with
XX peptide insertions, used in vaccines or for enzymatic reactions
XX
XX Example 8; Fig 5; 34pp; German.
XX
XX This invention describes a method for the production of a S-layer protein
XX (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
XX transformed with nucleic acid (II) encoding (I), linked to a signal
XX sequence (SS) that encodes a protein which causes at least one of (1)
XX integration of (I) into the external or cytoplasmic membranes and/or
XX (11) secretion of (I) into the periplasmic space or extracellular medium,
XX (b) culturing the cell to express (I) and (c) optionally recovering (I)
XX from the membranes, periplasmic space and/or extracellular medium.
XX Alternatively, a eukaryotic cell is used as host and then the SS, which
XX is optional, promotes integration into the cytoplasmic membrane or an
XX organelle and/or secretion into the extracellular medium. (I), and
XX derived structures, may include a wide variety of polypeptide inserts and
XX are useful as (1) vaccines or adjuvants (with immunogenic epitopes or
XX immunostimulants inserts such as cytokines) (11) as reactors (inserts
XX are enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a
XX 'molecular spinnerette' for production of PHB or luciferase for use as
XX molecular carrier molecule (streptavidin is inserted) for use in
XX universal carrier molecule (streptavidin is inserted) for use in
XX hybridisation and immuno assays, or for selective elimination of
XX cytokines, toxins etc. from body fluids (inserts are specific binding
XX epitopes). In this system, heterologous (I) do not form inclusion bodies
XX but rather monomolecular layers, and in eukaryotic cells they undergo
XX glycosylation. This sequence encodes a fusion gene constructed from the
XX Bacillus stearothermophilus sbas gene and the Bacteriophage fd Gene 3
XX signal sequence which is used in the method of the invention.
XX
XX SQ Sequence 3768 BP; 1320 A; 683 C; 733 G; 1032 T; 0 other;
XX
XX alignment_scores:
XX Quality: 248.50 Length: 898
XX Ratio: 0.597 Gaps: 44
XX Percent Similarity: 46.325 Percent Identity: 22.383
XX
XX alignment_block:
XX US-09-754-947-1 x AAX78246 ..
XX
XX Align seg 1/1 to: AAX78246 from: 1 to: 3768
XX
XX 55 AsnLeuProIleAspLysAspAlaLysProSerPheAlaAspSerGlnG 71
XX :::::  |||||
XX 937 GATGTTCTTATATCTACAGAGGAAACATCACTGCGATGCTCA.. 984
XX
XX 71 yGlnTrpTyrThrProPheIleAlaValGlnLysAlaGlnValIleL 88
XX |||||  |||||
XX 985 .....ACTGCATTTCGAAATAATACGAGATTAAGTAGTAGTTA 1024
XX
XX 88 ysgIyThrGlyAsnGlyPheGluProAsnGlyLysIleAspArgValSer 104
XX |||||  :::::
XX 1025 AAGGATTATAA.....GACAAATAATGCAAGAATTAAAGAAAGAT 1065
XX
XX 105 MetAlaSerLeuLeuValGlnAlaTyrLysLeuAspThrLysValAsnG 121
XX :::::  |||||
XX 1066 GCATTCACTTTCAGCTTCGAATGATGCTGATTACTCAAGTGTGG 1115
XX
XX 121 yThrProAlaThr..... 125
XX |||||  :::::
XX 1116 AACTATATGTAAACAACAACACTTCTGTAACCTTAGCAGCAGGTTCTTTCG 1165

```

126 LysPheLysAspLeuGlu 131
 1166 ACACGACGATCTTACAGTAGTATTGATTAAGTTGTTAGCAGCTGAA 1215
 132 ThrLeuAsn.....TrpGlyLys.. 137
 1216 ACTGTAACAGCGTCAGACCTTACTATTACAGATGTTGAAACGGAAGCG 1265
 138 GluLysAlaAsnIleLeuValGluLeuG 147
 1266 CATTCAGTAATTCGATCTACTCTGCTTCAATTAATTAATTAATTAATTA 1315
 147 LysIleSerValGlyThrGlyAspGlnTrpGlu.....ProLys 159
 1316 AAGAACGCTTACGTAAGTGTAAACATATTAACCTGCTATCAATATATGTT 1365
 160 LysThrValThr.....LysAlaGluAlaIleGlnPheIleAlaLysTh 174
 1366 AAMACATTAACGCTGTCACATGCGAAGCTTACGAGTTAGTG..... 1407
 174 RasPlySGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaVal 191
 1408TTCACTGCAAGCGATCGACCACTGTTGCTACCGCTC 1447
 191 aLThrThr..... 193
 1448 CTACTACTTTAGGTGTGACACTTATCTACTGTTCTTACACAAAT 1497
 193 193
 1498 GTTGGGCGTAATTCGCTGCTGTGTGAATGAAGCGAAGCTTATATTC 1547
 194GluLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
 1548 TGGTCTTCAATTCACAAACAGCTTCTCTACTAAGTAGACGAATCTACTT 1597
 209 ysgLysPheLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
 1598 TAGCTGATTAACCTTGTATTAGTTGAAGAAAGAAATCTGTACAGTTGTGCT 1647
 226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
 1648 TCTGAACCTTAATATATATGACAGCGCTAAATGTAACCTTGTAGTGCAAA 1697
 242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
 1698 ACGGACCTTAAAGAAATATCAATCTATCAATCAAAATTAAGAAAGCT 1747
 259 LysThrGlu.....ValAlaValGlySerLeuGluAlaLysThrIle 273
 1748 TCGAGTCGATTAAGGTATTGATTAAGCAGCTGTAAACGAAACATAT 1797
 274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnh 290
 1798 GAGTTCAAACCTCAGACTTAAGTCT.....CCTACAGTTATTAAGCT 1841
 290 eThrValLysAsp.....GluAsnGlyThrGluValLysSerProGluGlyI 306
 1842 AACGCTTAATAATGGCGACGCTGCATTAAGTAAGT.....GAAAGCTC 1885
 306 LeGluPheValThrProAlaIleGluLysIleAsnAlaLysGlyLeuIle 322
 1886 AAGAAATTTACTGTAAGTCTCAGAGAAATTTAAATACATTTAATGCTACA 1935
 323 ThrLeuAlaLysGlyThrSerThr..... 330
 1936 ACCGTTTCGGTAGCACATCAATACGCTCAAGTTGCTAGTAAAGC 1985
 331ThyV 332
 1986 GGGTGCAAACTATCTGCTTACAGCAAGTCACATCAATTCAGAGCTAGTG 2035

332 aLysAlaValTyrrLysLysAspGly.....LysValValAla 344
 2036 TTGAAGCGCTTACTGCTCAAGATGGACATACAAAGTGAAGTCTGCTGT 2085
 345 GluSerLysGlu..... 348
 2086 AACCAATTAATGAACGTAAACCAAGGTACAAATTAAGTGTGCTGAAGG 2135
 349ValLysValSerIleGluGluValAlaValAlaValAlaSerI 361
 2136 TGCACAGCTCCTGTTAAAGATGCTGCACAAATCAATTAATTAAGTACA 2185
 361 LeSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
 2186 ACTATATCATCATTTACAACTGAAGGTCAA...GACGTACACACACA 2232
 378 AspPheLysGlnAsnAsnLysValTyrrGluGlyAspAsnAlaTyrrValG 394
 2233 ACGGTT.....ACAAAGTATTCAAAGGTGATTC..... 2262
 394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
 2263TTAAAGACGCT..GATGCACTTACTACA..... 2289
 411 TyrGluSerLeuAsnThrGluValAlaValAlaValAspLysAlaThrGlyLys 427
 2289 2289
 428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal.... 442
 2290 CTTCAGAACGTTGATGACAGGTCAA.....AAATTCACATATCAATT 2330
 443LysAspSerLysGlyLysAlaLeuValSerHisThrTy 455
 2331 TAGCGAAGAAATTAATACTCTTAGTGCT...TCTTAGTGCGTGCAGAA 2377
 455 aGluIleGluAlaPheAlaGluLysAlaMetLysAspIleLysLeuGlu 471
 2378 TAACTGTCGCAAAATTAAACAAACGCGATGGGTGAT.....GCT 2418
 472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLy 488
 2419 GGTACTGGAACAACTGATACGTT..... 2442
 488 sAlaProValLeuAspGlnTyrrGlyLysGluPheThrAlaProValThrV 505
 2443 .GCTCTTAAGACAGATGCAATGATTAAGTAAACACCTGCTGTGTTACAT 2491
 505 aLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
 2492 TAACTGTCCTTGACAAATTAACGCAAAAGATGCGAAA...TTGCGTCTGTA 2538
 522 AlaLysTyrrValAsnArgGluLeuValLeuAsnAlaIleGln..... 536
 2539 GTAGATAAGTCTTCTACTGATGGAATTCGTATGAGCTGTAAGTAAAT 2588
 537GluAlaGlyAsnTyrrThrVal 544
 2589 TAAGAAAAAGATATTTAATTCGTTACACACGCTGAGACACACTGTAG 2638
 544 aLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
 2639 CTTCGTGAAGCTGCTGCTGCAAAAGATGTCGCA..... 2673
 561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
 2674AACGCTTCTGCTGATTCCTCA.....ACAAG 2699
 577 ThrGluLeuAsp.....LysTyrrValThrGluGluAsnGlnLysA 591
 2700 CACTCATTTGATTAACAATAGAGCTTATTAAGTAAATTCATGA.... 2745
 591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607

[illegible][illegible]

```

126 .....LysPheLysAspLeuGlu 131
1232 ACACTGACGATCTTAAACAGTAGTATTGATTGAAGTTAGTACACTGAA 1281
1332 ThrLeuAsn.....TyrGlyLys.. 137
1282 ACTGTAAACAGCTCGAACGTTACTATTACAGATGTTGAAACCTGGAAACG 1331
138 .....GluLysAlaAsnIleLeuValGluLeuG 147
1332 CATTCCAGTAATTCGATCTACTCTCGTCTCAATTAATTACTATTACTTTAA 1381
147 LylSerValGlyThrGlyAspGlnTyrPqlu.....ProLys 159
1382 AACAACTTAACTAGTAAGTAAACAAATATAAAGTCTGATCAATTAATGTT 1431
160 LysThrValThr.....LysAlaGluAlaLeuGlnPheLeuLysThr 174
1432 AAACACTTAATCGTTACATGCAAGCTTACGAGTTAGTG..... 1473
174 rAspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlav 191
1474 .....TTCACGTCAAAGCATCAGACCAACTGTTGTCACGCGTC 1513
191 alThrThr..... 193
1514 CTACTACTTAAAGTGTACAACTTATCTACTGTTCTCTTACAAACAAT 1563
193 ..... 193
1564 GTTTGGGTAATTTGGCTGTGTCGATGAAGCTGAACCTTATTATCC 1613
194 ...GlnLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
1614 TGGCTTTCATTCACAAACAAGCTTGTCTACTAAGTTAGACGAATCTACTT 1663
209 yAspLysValLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
1664 TAGCTGATTAAGTTGATTAGTTGAAAAAAGAAATCTGCTACAGCTGTGTC 1713
226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuVal 242
1714 TCTGAACCTAAATATATATGACAGCGCTAAATGCTTACTTATAGTGCAAA 1763
242 rSerAsnLeuAlaLysGlnThrTyrThrValAspValAsnLysValG 259
1764 ACCGCGACCTTAAAGAAATACAAATCAATCAAAATTAATAAAGGCT 1813
259 LysThrGlu.....ValAlaValGlySerLeuGlnLysThrIle 273
1814 TGAAGTCCGATTAAGCTATTGAATTAGCGACCTTAAACGAAACATAT 1863
274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnP 290
1864 GAGTTCAAAACACAGACTTAAGCTCT.....CTACAGTTATTAGCGT 1907
290 eThrValLysAsp...GluAsnGlyThrGluValValSerProGluGlyI 306
1908 AACGTTCTAAATAATGCGACGCTGATTTAAAGTAACT.....GAAAGTC 1951
306 LeuGluPheValThrProAlaAlaGluLysIleAsnAlaLysGlyLeuI 322
1952 AAGAAATTTACTGTGAAGTTCTCAGAGAAATTTAAATACATTAAATGCTACA 2001
323 ThrLeuAlaLysGlyThrSerThr..... 330
2002 ACCGTTTCGGGTAGCACATTCACATACGTCACATTGCTGTAGTAAAGC 2051
331 .....ThrV 332
2052 GGGTGAACAACTTATCTGCTTACAGCAAGTCACATCAATTCACAGCTAGTG 2101
332 alLysAlaValTyrLysLysAspGly.....LysValValAla 344

```

```

2102 TTGAAGCGGTACTGCTCAAGATGGAACATACAAAGTGAAGTGTCTGCT 2151
345 GluSerLysGlu..... 348
2152 AACCAATTAGAAACGTAAACCAAGGTACAAATTAAGTAGTGTGGTAAAG 2201
349 .....ValLysValSerAlaGluGluAlaAlaValAlaSerI 361
2202 TGCACACGCTCCTGTTAAAGATGCTGCACAAATGCAATTAATTAGCACTA 2251
361 LeSerAsnTyrThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
2252 ACTATATATATCAATTTACAACTGAAGGTCA...GACGTACACACACCA 2298
378 AspPheLysGlnAsnAsnLysValTyrGluGlyAspAsnAlaTyrValG 394
2299 ACGGTT.....ACAAAGTATTCAAGGTGATTTCT..... 2328
394 nValGluLeuLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
2329 .....TTAAAGACGCT..GATCGAGTTACTACA..... 2355
411 TyrGluSerLeuAsnThrGluValAlaValValAspLysAlaThrGlyLys 427
2355 ..... 2355
428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
2356 CTTACGAACGTTGATGACAGTCA.....AAATTCACATATCAATT 2396
443 .....LysAspSerLysGlyLysAlaLeuValSerIsthrV 455
2397 TAGCCAGACATTAATAAACTTCATGAGTGT...TCTTATGCGGTGCAAG 2443
455 alGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
2444 TAAGTGTGAGAAATTAACAACAACAGGATGGTAGAT.....GCT 2484
472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysValLys 488
2485 GGTACTGGACACACGTATCACTT..... 2508
488 sAlaProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrV 505
2509 ..GCTCTTAAGACAGATGCMAATGTAAAGTAAACACACTGCTGTGTACAT 2557
505 alLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
2558 TAAGTGTCTTGACAAATACACCAAAAGATGCGAAA...TTGCGTCTGGTA 2604
522 AlalysTyrValAsnArgGluLeuValLeuAsnAlaAlaGlyGln..... 536
2605 GTAGATAGTCTTCTACTGATGGAATTCGTGATGAGCTGTAATGTAAT 2654
537 .....GluAlaGlyAsnTyrThrValV 544
2655 TAAGGAAAAAGATATTTAATCTGTTACACAGCTGAGACACACTGTAG 2704
544 alLeuThrAlaLysSerGlyGluLysGluAlaLysAlaThrLeuAlaLeu 560
2705 CTCTGTGAAGCTGCTGCTGACAAAGATGTCAA..... 2739
561 GluLeuLysAlaProGlyAlaAlaPheSerLysPheGluValArgGlyLeuAs 577
2740 .....AACGCTTCGTGTCATTCCA.....ACAG 2765
577 PThrGluLeuAsp.....LysTyrValThrGluGluAsnGlnLysA 591
2766 CACTGCATTTGATCACTAAGAGCTTATGTAGTGAATTCATGA... 2811
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607

```

```

2812 .....ACTGATTACCGAGTAACCTGAGAACATCGTTT 2850
608 LysGIYAlaGluAlaIleuLysValThrThrAsnLysGluI 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2851 AAGATGACGACGATATGCGTAGCTGCTGATGACGATTAAGCG 2900
624 yLysGIYAlaSpAlaThrAspAlaGlnValThrValGlnAsnAsnSerV 641
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2901 TTCT.....ACAAATAAATTG 2917
641 AlIeThrValGIYGlnGlyAlaLysAlaGluThrThrTyLysValThr 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2918 TATTCACCTCCATCTCAGATTAATAAGCTGGTACAGTTACTGCTGATCA 2967
658 ValValLeuAspGly.....LysLe 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2968 ATT.....GACGGGTGAGAGATAAAGTAGTAACACATCTCAATA 3011
664 uIeThrThrHisSerPheLysValAlaSprThrAlaProThrAlaLysG 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3012 CATTAAT.....TCGTTCAAGACTGATCTGCGAATCCAAAG..... 3048
681 LysLeuAlaValaGluPheThrSerThrSerLysGluValAlaProAsn 697
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3049 .....TATCTTCAATCCACATTCGCTGACGCGTCACTTAC 3084
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValaAspGlyValPr 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3085 GTTGAC.....CGTTCTAATAAACAATTACATTCGATTCACGATTCGTTCC 3131
714 O...AlaThrThrAlaLysAlaThrAlaSerAsnValaGluPheValaSerA 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3132 AAACCAACCAATCACTCTTAAGAGCGACGAGCAATTCATTTACTAATT 3181
730 laasp...ThrAsnValaValaAlaGluAsnGlyThr.....ValGly 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3182 ACACCTTGTAAATGTAATATGAAATAAACAATACAAATTTGATTC 3231
743 AlaLysGlyAlaThr.....SerIleTyValLysAsnLeuThrValaVa 757
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3232 CACAAAGGTGAACACTGACGAGTTTACTCAATGATGAGTACGAGTTTC 3281
757 LysAspGlyLys.....GluGlnLysValaGluPheAspL 769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3282 AAAAGATTTCAAACTGCTACTGATATTGATGCAAAAGTTACATTCC... 3327
769 ysaIaValaGlnValaAlaValaSerIleLysGluAlaLysProAla 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3328 .....ATCACAGCTTCTGCTGCTACTGACGAAGTAAACCTGCT 3366
seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:AAx78245
seq_documentation_block:
ID AAX78245 standard; DNA: 4988 BP.
AC AAX78245;
AT
DT 24-AUG-1999 (first entry)
DE
XX B. stearothermophilus sbas/male fusion protein DNA sequence.
XX
XX sbas protein: S-layer protein; Gram-negative; prokaryotic host cell;
XX integration; cytoplasmic membrane; secretion; periplasmic space; toxin;
XX eukaryotic host cell; vaccine; adjuvant; immunogenic epitope; luciferase;
XX immunostimulant; cytokine; polyhydroxybutyrate; PHB synthase; body fluid;
XX molecular laser; universal carrier molecule; monomolecular layer;
XX fusion protein; male; ss.
OS Synthetic.
OS Bacillus stearothermophilus.
XX
XX DE19732829-A1.

```

```

PD 04-FEB-1999.
XX
XX PF 30-JUL-1997; 97DE-1032829.
XX
XX PR 30-JUL-1997; 97DE-1032829.
XX
XX PA (LUBIT/) LUBITZ W.
XX
XX PI Lubitz W, Resch S;
XX
XX DR WPI; 1999-122189/11.
XX
XX PT Producing S-layer proteins in Gram-negative bacteria or eukaryotes -
XX integrated into membranes or organelles or secreted into periplasma
XX or growth medium, and nucleic acid encoding S-layer proteins with
XX peptide insertions, used in vaccines or for enzymatic reactions
XX
XX PS Example 7; Fig 4; 34pp; German.
XX
XX CC This invention describes a method for the production of a S-layer protein
XX CC (I) which comprises (a) preparing a Gram-negative prokaryotic host cell
XX CC transformed with nucleic acid (II) encoding (I), linked to a signal
XX CC sequence (SS) that encodes a protein which causes at least one of (1)
XX CC integration of (I) into the external or cytoplasmic membranes and/or
XX CC (II) secretion of (I) into the periplasmic space or extracellular medium,
XX CC (b) culturing the cell to express (I) and (c) optionally recovering (I)
XX CC from the membranes, periplasmic space and/or extracellular medium.
XX CC Alternatively, a eukaryotic cell is used as host and then the SS, which
XX CC is optional, promotes integration into the cytoplasmic membrane or an
XX CC organelle and/or secretion into the extracellular medium. (I), and
XX CC derived structures, may include a wide variety of polypeptide inserts and
XX CC are useful as (1) vaccines or adjuvants (with immunogenic epitopes or
XX CC immunostimulants inserts such as cytokines) (II) as reactants (inserts
XX CC are enzymes, e.g. polyhydroxybutyrate (PHB) synthase for use as a
XX CC 'molecular spinnerette' for production of PHB or luciferase for use as
XX CC universal carrier molecule (streptavidin is inserted) for use in
XX CC hybridisation and immuno assays, or for selective elimination of
XX CC cytokines, toxins etc. from body fluids (inserts are specific binding
XX CC epitopes). In this system, heterologous (I) do not form inclusion bodies
XX CC but rather monomolecular layers, and in eukaryotic cells they undergo
XX CC glycosylation. This sequence encodes a novel fusion protein constructed
XX CC from the Bacillus stearothermophilus sbas gene and the male gene signal
XX CC sequence which is used to illustrate the method of the invention.
XX
XX SQ Sequence 4988 BP; 1678 A; 985 C; 1052 G; 1273 T; 0 other;

alignment_scores:
Quality: 248.50 Length: 898
Ratio: 0.597 Gaps: 44
Percent Similarity: 46.325 Percent Identity: 22.383

alignment_block:
US-09-754-947-1 x AAX78245 ..

Align seg 1/1 to: AAX78245 from: 1 to: 4988

55 AsnLeuProIleAspLysAspAlaIleuProSerPheAlaAspSerGlnI 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2131 GATGTTTCTTATCTACAGACGAGAAAACATCTACCTGCGATGCTTCA.. 2178
71 yGlnTrpTyThrProPheIleAlaIleuValaGluLysAlaGlyValIleI 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2179 .....ACTCATTCGAAATAATTAATACGAGATTAATAGTAGTAGTTA 2218
88 ysgLyThrGIYAsnGIYPheGluProAsnGlyLysIleAspArgValSer 104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2219 AAGGATTTAA.....GACAAAAATGGCAAAAGAAATTTAAAGAAAT 2239
105 MetAlaSerLeuLeuValaGluAlaTyLysLeuAspThrLysValaAsnGI 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2260 GCATTCACCTTCAAGCTTCGAATGATGCTGATGCTACTCAAGTGTGG 2309

```

121 yThrProAlaThr..... 125
 |||||
 2310 AACTAATGTAAACAACAACACTTCTGTAACCTTAGCAGCAGGACTTTGCG 2359
 126 LysPheLysAspLeuGlu 131
 |||||
 2360 ACACGTGACGACTCTTAAACAGTAGATTTGATTTGATTGTTACACCTGAA 2409
 132 ThrLeuAsn..... TrpGlyLys.. 137
 |||||
 2410 ACTGTAAACGCTCGAACCTTACTATTACAGATGTTGAACCTGGAAACG 2459
 138 GluLysAlaAsnIleLeuValGluLeuG 147
 |||||
 2460 CATTCAGTAATGTGACTCTACTCTGCTTCTACAAATTACTATTAGCTTAA 2509
 147 LysIleSerValGlyThrGlyAspGlnTrpGlu..... ProlLys 159
 |||||
 2510 AAGAAGCGTTAGTAACTGGTAAACAATATMAACTGCTATCAATATGCTT 2559
 160 LysThrValThr..... LysAlaGluAlaAlaGlnPheIleAlaLysThr 174
 |||||
 2560 AAAACCTTAACGTGTTACAAATGACAGAGCTTACGAGTTAGTG..... 2601
 174 rAspLysGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaV 191
 |||||
 2602 TTCACTGCAACGACATACACCAACTGTTGCTACCGCTC 2641
 191 aLThrThr..... 193
 |||||
 2642 CTACTACTTCTTAGGTGTACACTTATCTACTGTTCTTTCACAACAAT 2691
 193 193
 2692 GTTTGGCGTAATTTGGCTGTGTGTGATGAAGTGAAGCTGAACCTATTATTC 2741
 194 GluLysValGluValLysPheSerLysAlaValGluLysLeuThrL 209
 |||||
 2742 TCGCTCTTCAATTCACAAACACGCTTTCCTACTAAGTTAGACGAATCTACTT 2791
 209 yGluAspIleLysValThrAsnLysAlaAsnAsnAspLysValLeuVal 225
 |||||
 2792 TAGCTGATTAACCTTTGATTAGTTGAAAAAGAAATGTGTACACTGTGTGCT 2841
 226 LysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeuTy 242
 |||||
 2842 TCTGAACCTAAATTAATATGACAGCCCTAAATGCTAATCTTAGTCCCAA 2891
 242 rSerAsnLeuAlaAlaLysGlnThrTyThrValAspValAsnLysValG 259
 |||||
 2892 ACCGGACCTTAAAGAAATACAAATCTCAAAATTAAAGGCT 2941
 259 LysThrThrGlu..... ValAlaValGlySerLeuGluAlaLysThrIle 273
 |||||
 2942 TCGAAGCCGATTAACGTAATTAATTAAGCAGCTTAAACGAGAAACATAT 2991
 274 GluMetAlaAspGlnThrValAlaAlaAspGluProThrAlaLeuGlnPh 290
 |||||
 2992 GAGTTCAAAACACAGACTTACTCTCT..... CCTACAGTTATATAGCCT 3035
 290 eThrValLysAsp..... GluAsnGlyThrGluValValSerProGluGlyI 306
 |||||
 3036 AACGTCCTAAAAATGGCAGCGCTGATTTAAAGTAACT..... GAAAGCTC 3079
 306 LeuPheValThrProAlaAlaGluLysIleAsnAlaLysGlyLysIle 322
 |||||
 3080 AAGAATTTACTGTGAAGTTCTCAGAGAAATTTAAATCAATTAATATCTCA 3129
 323 ThrLeuAlaLysGlyThrSerThr..... 330
 |||||
 3130 ACCGTTTCGGGTAGCACAATCAATACGCTCAAGTTGCTGTAGTAAAGC 3179

331 ThrV 332
 3180 GGGTCAAACTTATCTGCTCTTACAGACAAGTGCATCATTCACAGCTAGTG 3229
 332 aLysAlaValLysLysAspGly..... LysValValAla 344
 |||||
 3230 TTGAAGCGGTACTGCTGCAAGATGGACATACAAAGTGAAGTTGCTCT 3279
 345 GluSerLysGlu..... 348
 |||||
 3280 AACCAATTAAGAACGTACCAAGGTACAAATTAAGTAGTCTTCGTAAAG 3329
 349 ValLysValSerAlaGluGlyAlaAlaValAlaSerI 361
 |||||
 3330 TGCACACGCTCCTGTTAAAGATGCTGCAAAATGCAATCTTAGCAACTA 3379
 361 IeSerAsnTrpThrValAlaGluGlnAsnLysAlaAspPheThrSerLys 377
 |||||
 3380 ACTATATCTATACATTTACAACTGAAGGTCAA... GACGTACACAGCACA 3426
 378 AspPheLysGlnAsnAsnLysValTyroGluLysAspAsnAlaTyValG 394
 |||||
 3427 ACGGTT..... ACAAACTATTCAAAGCTGATTCCT..... 3456
 394 nValGluLysAspGlnPheAsnAlaValThrThrGlyLysValGluT 411
 |||||
 3457 TTAAAGACGCT... GATGCAGTTACTACA..... 3483
 411 TyrGluSerLeuAsnThrGluValAlaValAlaAspLysAlaThrGlyLys 427
 3483 3483
 428 ValThrValLeuSerAlaGlyLysAlaProValLysValThrVal..... 442
 |||||
 3484 CTTACGAACGTTGATGACAGCTCA..... AATTCACATTCACAT 3524
 443 LysAspSerLysGlyLysAlaLeuValSerHisThrV 455
 |||||
 3525 TAGCGAAGAAATTAACAAACGATGAGTGT... TCTTATGTTGGCAAG 3571
 455 aGluIleGluAlaPheAlaGlnLysAlaMetLysAspIleLysLeuGlu 471
 |||||
 3572 TAACGTGCGAAGAAATTAACAAACGATGAGTGTAT..... GCT 3612
 472 LysThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysVal 488
 |||||
 3613 GSTACTGGAACAACGCTATCAAGT..... 3636
 488 sAlaProValLeuAspGlnTyroGlyLysGluPheThrAlaProValThrV 505
 |||||
 3637 .GCTCTTAAGACGATGCAAAATGTTAAAGTAAACGCTGCTGTGTATCAT 3685
 505 aLysValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuGlu 521
 |||||
 3686 TAACGTGCTTGTACATTAACGACAAAGATGCGAA... TTGCGTCTGTA 3732
 522 AlaLysTyroValAsnArgLysLeuValLeuAsnAlaAlaGlyGln..... 536
 |||||
 3733 GTAGATAAGCTGTCTACATGATGGAATTCGTATGAGCTGTAATGTAA 3782
 537 GluAlaGlyAsnTyroValV 544
 |||||
 3783 TAAGAAAAAGATATTTAAATTCGTTACAAACGCTGACGACACTGTAG 3832
 544 aLeuThrAlaLysSerGlyLysGluAlaLysAlaThrLeuAlaLeu 560
 |||||
 3833 CTCTGTGAAAAGCTGCTGACAAAGATGCTCA..... 3867
 561 GluLeuLysAlaProGlyAlaPheSerLysPheGluValArgGlyLeuAs 577
 |||||
 3868 AACGCTTCGCTGCAATCCCA..... ACAG 3893
 577 PThrGluLeuAsp..... LysTyroValThrGluGluAsnGlnLysA 591

```

3894 CACGCAATTCATCACTAGACCTTATTAGTGAATTCATGAA... 3939
591 snAlaMetThrValSerValLeuProValAspAlaAsnGlyLeuValLeu 607
3940 .....ACTGATTACGGGAAGTTAAACCTGACAGAACATCGTTGTT 3978
608 LysGlyAlaGluAlaAlaGluLeuLysValThrThrAsnLysGluG 624
3979 AAAGATGACAGAGTAATGCGTACTGCTAGCTGTAACAGCATTAAGACG 4028
624 yLysGluValAspAlaThrAspAlaGluValThrValGlnAsnSerV 641
4029 TTCT.....ACAATAAATTG 4045
641 alIleThrValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThr 657
4046 TATTCACGTCATCTCAAGATTAAAGCTGCTAGCTTACTCTTAACA 4095
658 ValValLeuAspGly.....LysLe 664
4096 ATT.....GACGTCGAGAGATTAAGTAGTACACACATCTTAATA 4139
664 uIleThrThrIleSerPheLysValValAspThrAlaProThrAlaLysG 681
4140 CATTACT.....TCGTTCAAGACTGATCTGCCAATCCACG..... 4176
681 yLeuAlaValAlaGluPheThrSerThrSerLeuLysGluValAlaProAsn 697
4177 .....TTATCTTCATCAGCATTTGCGACGGTGACATTAC 4212
698 AlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerValAspGlyValPr 714
4213 GTTGAC.....CGTTCTAAACAATTACATTGATTCAGCGATTGAGTTCC 4259
714 C...AlaThrThrAlaLysAlaThrAlaSerAsnValGluPheValSerA 730
4260 AAACCCAACTCACTCTTAAGAGCGTACGGAACCTCATTTACTTAAT 4309
730 laAsp...ThrAsnValValAlaGluAsnGlyThr.....ValGly 742
4310 ACACCTTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATC 4359
743 AlaLysGlyAlaThr.....SerIleTyrValLysAsnLeuThrValVa 757
4360 CACAAGGCTGAACACTGACGAGTTTACTCAATATGAGTTAGCAGTTTC 4409
757 lLysAspGlyLys.....GluGlnLysValGluPheAspL 769
4410 AAAAGATTTCAAACTGCTACTGATATTGATGCAAGCTTACATTTC.... 4455
769 ysaAlaValAlaValAlaValSerIleLysGluAlaLysProAla 783
4456 .....ATCACAGGTTCTGTTGCTACTGACGAAAGTAAACCTCTCT 4494
seq.name: /STDS2/9c9data/geneseq/geneseq/NA1992.DAT:AAQ29471
seq_documentation_block:
ID AAQ29471 standard; DNA; 6744 BP.
XX
AC AAQ29471:
XX
DT 04-MAR-1993 (first entry)
XX
DE Extracellular factor related protein gene.
XX
KM EF*: detection; prevention; screening; diagnostic; ss.
XX
OS Streptococcus suis type II (non-pathogenic).
XX
FH Key location/Qualifiers
FT CDS 361..5826
FT /*tag= a

```

```

FT sig_peptide 361..498
FT /*tag= b
FT mat_peptide 499..5826
FT /*tag= c
FT promoter 66..71
FT /*tag= d
FT promoter /note= "-35 region"
FT 89..94
FT /*tag= e
FT promoter /note= "-10 region"
FT 153..158
FT /*tag= f
FT promoter /note= "-35 region"
FT 176..181
FT /*tag= g
FT RBS /note= "-10 region"
FT 350..356
FT /*tag= h
FT repeat_unit /note= "ribosome binding site"
FT 2869
FT /*tag= i
FT repeat_unit /note= "start of repetitive unit R1"
FT 3097
FT /*tag= j
FT repeat_unit /note= "start of repetitive unit R2"
FT 3292
FT /*tag= k
FT repeat_unit /note= "start of repetitive unit R3"
FT 3520
FT /*tag= l
FT repeat_unit /note= "start of repetitive unit R4"
FT 4087
FT /*tag= m
FT repeat_unit /note= "start of repetitive unit R5"
FT 4381
FT /*tag= n
FT repeat_unit /note= "start of repetitive unit R6"
FT 4609
FT /*tag= o
FT repeat_unit /note= "start of repetitive unit R7"
FT 4837
FT /*tag= p
FT repeat_unit /note= "start of repetitive unit R8"
FT 5065
FT /*tag= q
FT repeat_unit /note= "start of repetitive unit R9"
FT 5293
FT /*tag= r
FT repeat_unit /note= "start of repetitive unit R10"
FT 5521
FT /*tag= s
FT repeat_unit /note= "start of repetitive unit R11"
FT 6554..6566
FT /*tag= t
FT misc_feature /note= "dyad symmetry region 1a"
FT 6571..6583
FT /*tag= u
FT misc_feature /note= "dyad symmetry region 1b"
FT 6611..6625
FT /*tag= v
FT misc_feature /note= "dyad symmetry region 2a"
FT 6631..6644
FT /*tag= w
FT /*note= "dyad symmetry region 2b"
XX
PN W09216630-A.
XX
PD 01-OCT-1992.
XX
PF 19-MAR-1992;
XX
PR 21-MAR-1991;
XX
92MO-NL00054.
91NL-0000510.

```

136 GlyLysGluLysAlaAsnIleLeuValGluLeuGlyIleSerValGly.. 151

3520	CAGAAAGATGCTTAACACAAGAATTGCCAANAAGACAGATCAAGCTTAACAA	3568
152ThrGlyAspIleThrGluProLysL	160
3570	AGCGATTGAAGCAATCCAAACTTGACTGCATGATAAGAAAGAAATCAGTGA	3619
160	ySThrValThrLysAlaGluAlaIleGlnPheIleAlaLysThrAspLys	176
3620	AGATTCACGTGGAMAGGGCGCTTAAGGTAGACAAACCGCATTTGATGATAA	3669
177	GlnPheGlyThrGluAlaIleLysValGluSerAlaLysAlaValaThr....	192
3670	GCATCACTCCAGATGCAAGTGTAAAGTAGAAGGACAAAGGTGTATGCAGC	3719
193ThrGlnLysValGluValLys.....	199
3720	TATCATTTGATTACTGCGCACGAGATGCTAAAGTGTCAATTCGTGCTGA	3769
200PheSer	201
3770	AGTTGGCAGATGAAAATCAGAAAGCTCGAAAGATAACAGAACAGAAAGCAA	3819
202	LysAlaValaGlu.....LysLeuThrLysGluAspIleLysValaThrAs	216
3820	AAGCTATGATGATCGCTCAACTGTGACTATATGAGGAGAAAGCAATCGCTTAA	3869
216	nlysAlaAsnAsnAspLysValaLeuValLysGluValaThrLeuSerGluA	233
3870	GAAAGCTCTTCAAGATGTTGTGATTAAGAAAGAAACACAGCTTGAAAGCG	3919
233	splysArgSerAlaThrValGluLeuTrf.....	242
3920	CAGCTAGGTGACCACAAATGATGATTCATGAAAGCTATTACTACAGAAAAA	3969
243SerAsnLeuAlaLysGlnThrTrf.....	252
3970	GCAGAAAGCGGGCACTGCTGGCAAAAGACCTTGACACAGACACAGCTAA	4019
253ValAspValaSnLys.....ValGlyLysT	261
4020	AGAGCTACAGATGCAGTTGATTGGCTAAGGATTAAGATAAATTAGGTAAAG	4069
261	hrgLyuAlaValaGlySerLeuGluAlaLysThrIle.....GluMet	275
4070	AAGCAATCCGACACAGAAAGAAAGACGTACTAAATAGTAGAGAAACTT	4119
276	AlaAspGlnThr.....ValValaLlaAspGluProThrAlaLeuGlnPh	290
4120	GCAGAAAGATACCGCAAAAGCTATGACGACATTCCA.....	4155
290	eThrValLysAspGluAsnGly.....ThrGluV	300
4156	.AACTTGTCAGATGAAGATAACCAAGCGGAATTAAAAAGTAACTGACG	4204
300	aValaSerPro.....GluGlyIleGluPheValaThr	310
4205	CTGTGGCAAAAATTYAGCAACCATGCGTGCACAAAGCAGATTAACGTAACG	4254
311	ProAlaIleAgLuLysIleAsnAlaLysGlyGluIleThrIleuAlaLysGl	327
4255	CAGAAGACGAAAAGCTCAAMGCCCTAGCAGCTTTGAAAAAGCTTAAGA	4304
327	yThrSerThrThrValLys.....AlaValaTryL	337
4305	AACACAGAAAATTTGGCAGATTAAGCTGCATGTATGTTGACTTACTGTCG	4354
337	ysLysAspGlyLysValaValaLagIuserylsgLuvAlaLys.....	350
4355	TGAAGATGTCAGCTTAAGCTTAAACAAGATGCTAAAGACACAGATT	4404
351ValaSerAlaGluGluValaValaValaAlaSerIleSerAsnTr	364
4405	GCTAAAGATGACGCGCTGCTTAAGAAAGCAATTTGCAACCAATTCCTT	4454

```

364 pThvAlaIaGlulnAsnLysAlaAspPheThr.....SerL 377
    ||| ||||| ||| |||||
4455 GACGACGCGCAG.....AAGAAACCTTCACGATCGCGTAGATGCAG 4498
377 ysAspPheLysGlnAsn.....Lys 384
    ||| ||| ||| |||
4499 AAGTAGCCAAAGCTAACGACGCAATTTCAGTCGACACCGCCAGCAT 4548
385 ValTyrGlulnLysAspAlaTyrValGlnValGluLeuLysAspGlnPh 401
    ||| ::::: ||| ||| ::::: |||
4549 GTTCAGAAAGAGAGAGATGCAGGTGTCACAGCATTCGACAGAAAGATGTTCT 4598
401 eAsnAlaVal.....ThrThrGlyLysValGluTyrGluSerLeuA 415
    ::::: ||| ::::: ||| ::::: |||
4599 TGACGCGAGCTAAACAGATGCTAAGAAATTAAGATTGCTAAAGATGCAGCG 4648
415 snThrGluValAlaVal..... 420
    ::::: ||| ::::: |||
4649 CTGCTAAAGAAAGCAATTGGCTCCAAATCCAAACTGCACAGACCGACAGAG 4698
421 .....ValAspLysAlaThrGlyLysValTyr...Va 430
    ||||| ::::: ||| ::::: |||
4699 AAACCTTACCAGATCGCGTAGATGCAGAAAGTAGCCAAAGCTAAAGACGC 4748
430 lLeuSerAlaGlyLysAlaProValLysValThrValLysAspSerLysG 447
    ::::: ||| ::::: ||| ::::: |||
4749 AATTTCAGCTGCCAACCAAGCCACGACGATGTTCAAAAAGAGAGAGATGCAG 4798
447 lLysAlaLeuValSerHisThrValGluIleGluAlaPheAlaGlnLys 463
    || ||| ::::: ||| ::::: |||
4799 GTGTTCGACCATTCGACAGAGATGTT...CTTGACGCGAGCTAAACAGAT 4845
464 AlameLysAspIleLysLeuGluLysThrAsnValAlaLeuSerThrLy 480
    ||| ::::: ||| ||| ::::: |||
4846 GCTAAGATAGATTGCTTAAAGAA...TCGACGCTGCTAAGTCAAGCAT 4892
480 sAspValThrAspLeuLysValLysAlaProValLeuAspGlnTyrGlyL 497
    ||| ::::: ||| ||| ::::: |||
4893 TGACGCGCAT.....CCAACTTCAGACAGATGCAGAGA 4924
497 ysgLupPheThr.....AlaProValThrValLys 506
    ||||| ::::: ||| ::::: |||
4925 AGGAATCGCTACAGAAAGCAGTGTGATGCTAAAGCTGCGACAGAT 4974
507 ValLeuAspLysAspGlyLysGluLeuLysGluGlnLysLeuAlaLys 523
    ::::: ||| ::::: ||| ::::: |||
4975 GCATTTGATGCTTCAACAAGTCCAGTCGAAAGCGCAATCGGACAGACAA 5024
523 sTyrValAsn.....ArgGluLeuValLeuAsnAlaIaGlyGlnGluA 538
    ||| ::::: ||| ||| ::::: |||
5025 AGCGCTAGGTTCAATCGCCCAAGATGTTCTTGACGCGACGAAACAAAGATG 5074
538 lAglYAsnTyrThrValValLeuThrAlaLysSerGlyGlnLysLysLys 554
    ||| ||| ||| |||
5075 CTAAGAAC.....AAGATTGCCAAAGAAAGTTGCC 5103
555 LysAlaThrLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPh 571
    ||| ||| ::::: |||
5104 GCAGCTAAGAGACATGATGCCAATCCG..... 5133
571 eGluValArgLysLeuAspThrGluLeuAspLysTyrValThrGluGluA 588
    ::::: ||| ::::: ||| ::::: |||
5134 .AACTTATCAGATGCAGAGAAAGAGCTTCTAAGAAAGCGGATGATGCAG 5182
588 snGlnLys.....AsnAlaMetThrValSerValLeuProValAsp 601
    ::::: ||| ::::: ||| ::::: |||
5183 ATGCTAAGCTACGACAGATGCAATGATGCTTCAACAAGTCCAGTCCAA 5232
602 AlaAsnGlyLeuValLeuLysGlyValA..... 610
    ||| ::::: ||| ||| ::::: |||
5233 GCGCAATCGCAGAGACAAAGCGCTAGTTCAATCGCCCAAGATGTTCT 5282

```

```

611 .GluAlaIaGluLeuLysValThrThrThrAsnLysGluGlyLysGluV 627
    ::::: ||||| ||| ::::: |||||
5283 TGACGACGCG.....AAACAAGATGCTAAGCAATTAAGATTGCTTAAAGAT 5326
627 alAspAlaThrAspAlaGlnValThrValGlnAsnAsnSerValIleThr 643
    ||||| ::::: ||| ::::: ||| ::::: |||
5327 CCGACGCTGCTAAGTCAAGCATTGACCGGAATCCAACTTGACAGATGCA 5376
644 ValGlyGlnGlyAlaLysAlaGlyGluThrTyrLysValThrValLe 660
    ::::: |||||
5377 GAGAAGCAATCAGCTAAG..... 5394
660 uAspGlyLysLeuIleThrThrHisSerPheLysValValAspThrAlaP 677
    ||| ::::: |||||
5395 .....AAAGCGGTAGAT..... 5406
677 roThrAlaLysGlyLeuAlaValGluPheThrSerThrSerLeuLysGlu 693
5406 ..... 5406
694 ValAlaProAsnAlaAspLeuLysAlaAlaLeuLeuAsnIleLeuSerVa 710
    ||||| |||||
5407 .....GCAGATGCTTAAAGCTGCG.....AC 5426
710 lAspGlyValProAlaThrThrAlaLysAlaThrAlaSerAsnValGlu. 726
    ||| ::::: ||| ::::: ||| ::::: |||
5427 AGATGCAATTGATGCTTCAACAAGTCCAGTCGAAGCCCAATCGGACAGAG 5476
727 .....PheValSerAlaAsp..... 731
5477 ACAAGGCGTAGGCGCCATCGCCAAAGACATTTGATGCCGCCAAACAA 5526
732 .....ThrAsnValValAlaGluAsnGlyThrValGluAlaLysGlyAl 746
    ||| ::::: ||| ||| ::::: |||
5527 GATGCTAAGACAGATGCTTAAAGAG.....GCAGATCCGC 5564
746 aThrSerIleTyrValLysAsnLeuThrValValLysAspGlyLysGlu. 762
    ||| ::::: ||| ||| ::::: |||
5565 TAACTCAGTCTAATGCTCCAACTCCAGTTCACAGATGCACGCTAAGAG 5614
763 .GlnLysValGluPheAspLysAlaValGlnValAlaVal 775
    ||| ||| ::::: ||| ||| ::::: |||
5615 CCGCTAATCTGAATATGATTAAGCTGTTGAGAAAGCAT 5655
seq_name: /SID2/gcgcdata/geneseq/geneseqn/NA1996.DAT.AAT41476
seq_documentation_block:
ID AAT41476 standard; DNA; 7291 BP.
XX
AC AAT41476;
XX
DT 15-JAN-1997 (first entry)
XX
DE Haemophilus adhesion protein HA2 gene.
XX
KW Haemophilus adhesion protein; HA2; hsf protein; vaccine; ss.
OS Haemophilus influenzae type b strain C54.
XX
FH key Location/Qualifiers
FT misc_difference 4
FT /*tag= a
FT /note= "base 4 is given as n in the specification"
FT misc_difference 57
FT /*tag= b
FT /note= "base 57 is given as n in the specification"
FT misc_difference 59
FT /*tag= c
FT /note= "base 59 is given as n in the specification"
FT CDS
FT /*tag= d
XX
PN W09630519-A1.

```

XX 03-OCT-1996.
PD 22-MAR-1996; 96WO-US04031.
XX 24-MAR-1995; 95US-0409995.
PR (UYSL-) UNIV ST LOUIS.
PA (UNIM) UNIV WASHINGTON.
XX Barenkamp SJ, St Geme JW;
XX WPI. 1996-455364/45.
DR P-PSDB; AAR9393.
XX
PT Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
XX vaccines against H. influenzae infection.
XX
PS Claim 8; Page 55-65; 120pp; English.
XX
XX A genomic DNA clone (AAT41476) codes for the Haemophilus adhesion
CC protein HA2 (AAR9393) of H. influenzae type b strain C54. It was
CC isolated from a 8.3 kb XbaI fragment of plasmid pDC400, a pUC19
CC derivative contg. the C54 surface fibrin locus. HA2 is associated
CC with the formation of surface fibrils involved in adhesion to host
CC cells. The isolated DNA can be incorporated into a vector and used
CC to direct expression of recombinant HA2 in transformed host cells
CC for use in vaccines against H. influenzae infection.
XX
XX Sequence 7291 BP; 2508 A; 1316 C; 1758 G; 1706 T; 3 other;

alignment_scores:
Quality: 240.50 Length: 850
Ratio: 0.584 Gaps: 41
Percent Similarity: 48.471 Percent Identity: 22.706

alignment_block:
US-09-754-947-1 x AAT41476 ..

Align seg 1/1 to: AAT41476 from: 1 to: 7291

23 GLULYSGLYALVALYGLYSNAPLYSGLYMELPHEGLUPROGLYLY 39
||||| :
397 GAAAAAGAGTTACAGAAATTCAAATGGGAAATATTTTCGACAAATAA 446
39 sGLULYTHRRARGALAGLUALALATHRMELALAGLILLEULASNL 56
| :
447 ACGAGTACTAAAGACCGGAGCAATCACCCTCAAGCCGCCGAC..AAC 493
56 euPPIleAsplysAspAla.....LysProSerPheAla 67
|| ||| :
494 TGAATAATCAACAAAACCCGATGAAGCACCACCATGCCAGTTCACCC 543
68 AspSerGInGlyInTrpThrProPheIleAla..AlaValGluLy 83
||| :
544 TACTGCGTGAAGAAAGACCTCACAGATCGACCAGTGTTCAGAAAGAAA 593
83 sAlaGlyVal.....IleLysGlyThrGly 92
| :
594 ATTATCGTTTGGCGCAACGCGCATTAAGTATTTACCACTGATGCCAA 643
92 sngLysPheGlu.....ProAsnGlyLysIle..... 100
644 ATGCGCTGAATTTGGCAAAACAGGTAACGCAATGTTTCATTTGAATGT 693
101AspArgValSerMetAlaSerLeuLeuValGlu 111
694 TTGGATTCAACTTTCCTGATGCGGTAAACGATACAGGTGTGTAAATTC 743
111 uAlaLysLysLeuAspThrLysValAsnGlyThrProAlaThrLysPheL 128
:
744 ATCAAGTTTACACCTAATGATGTGAAAAAACAAGAGCTGCACTGTTA 793

128 yAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleLeuVal 144
||| :
794 AAGATGTTTAAATGACGGTTGCAACATTAAAGGCTAAACT..... 837
145 GluLeuGlyIleSerValGlyThrLysPglInTrpGluProLysLysTh 161
838GCTGAGCTAATGTTGAGAGCTGTGATTT 866
161 rValThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAspLysGlnP 178
867 AGTGTCCTTATTAATGTTGAATTTATTAACAGCCGATTAACACGCC 916
178 heGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrThrGln 194
917 TTGATGTGTGTTAATACAGCTAAAGAAACGGTAAACA..... 954
195 LysValGluValLysPheSerLysAlaValGluLysLeuThrLysGluAs 211
955 ...ACCGAAGTGAAATTCAACCCGAAACCTCTGTATCAAGAAAAAGA 1001
211 pIleLysVal...ThrAsnLysAlaAsnAsp.....LysValLeuV 225
1002 CGGTAACTTATTACTGCAAAAGACATTAACGACCAATTAAGTTACAA 1051
225 aLysGluValThrLeuSerGluAspLysArgSerAlaThrValGluLeu 241
1052 GTAAACACGGCCAGCTGATATACAGATGAGGTAATGGCTTACACTGCA 1101
242 TyrSerAsnLeuAlaAlaLysGlnThrThrValAspValAsnLysVal 258
1102 AAAGCTGTGATTGATGCT.....GTGAACAAAGC 1130
258 lGly.....LysThrGluValAla..... 264
1131 TGCTTGAGAGCTTAACACACTGCTAATGCTAAATGCGCAGCTTCG 1180
265ValGlySerLeuGluAlaLysThrIleGluMetAlaAspGlnThr 279
1181 CAAGTGTGCGTCACGACCAATGTAACCTTTGAAGTGGCGATGCTACA 1230
280 ValValAlaAspGluProThrAlaLeuGlnPheThrLysAspGluAs 296
1231 ACAGCGTCA.....GTAACCTAAAGATACCTAA 1256
296 nGlyThrGluValAlaSerPro.....GluGlyIleG 307
1257 CGGCAATGGCATCACTGTTAAGTACGACGCAAAAGTTGCGACGGCTTGA 1306
307 LuPheValThrProAlaAlaGluLysIleAsnAlaLysGlyGluIleThr 323
1307 AATTT.....GATAGCGATAAAAAATC..... 1329
324 LeuAlaLysGlyThrSerThrThrValLysAlaValTyrLysLysAspG 340
1330 GTTGCGATTAACGCGCAGCTTACTGTGTACAGCT.....GG 1364
340 yLysValAlaLagIleSerLysGlu.....ValLysValSerAlaG 354
1365 TAAAGTACTGAATTTGCTAAAGAAAGATGACAGAAAAAACTTGTATG 1414
354 LuGlyAlaAlaValAlaSerIleSerAsn.....TyrThrValAlaGlu 368
1415 CAGCGCATTTGTGAACACCTTAGGTATCTAAGTTGGAAAGCAAAAGCT 1464
369 GluAsnLysAlaAspPheThrSerLysAspPheLysGluAsnAsnLysVal 385
1465 GAGCGTGATGATGATGCGCTTGAGGGGATTTCAAAAGACCAAGAGT 1514
385 lTyrGluLysAspAsnAlaTyrValGlnValGluLeuLysAspGlnPheA 402
1515 CAAAGCAGGCAACACGTA.....ACCTTTA 1540

```

402 snAlaValThrGlyLysValGluTyrGluSerLeuAsnThrGluVal 418
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1541 AAGCGGCGAAGACTTAAAGCTAAAGAGATGGCGAAGCTTACTTAT 1590
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
419 AlaValAlaAspLysAlaThrGlyLysValThrVal 430
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1591 TCACGCAAGATGCTTTACGGGTTTACGAGCATTAAGTGGTAC 1640
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
431 LeuSerAlaGlyLysAlaProValLysValThrValLysAsp 444
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1641 AACATATGGCGGAATGATGGCAAAACCGTCATCAACAAGCGTTTAA 1690
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
445 SerLysGlyLysAlaLeuValSerHisThrValGlu 456
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1691 CCATCACCAGCCAGCATATGGCGGTACAGACATCAACACCATCAGC 1740
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
457 IleGluAlaPheAlaGlnLysAla...MetLysAspIleLysGluGly 472
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1741 GTAACCAAGATGGCATTTAAGCAGTAAATTAAGCTATT 1779
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 sThrAsnValAlaLeuSerThrLysAspValThrAspLeuLysVal 489
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1780 ACTAATGTCGAGTGGTTAAGAGCTTAAGCAT 1815
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
489 lAlaProValLeuAspGlnTyrGlyLysGluPheThrAlaProValThrVal 505
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1816 GCGAATTTT 1824
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
506 LysValLeuAspLysAspGlyLysGluLeuLysGluGln 518
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1825 GATGTTTAAATTAAGCTGCAAGCTTAATTAATAGACAGTTGAAGATGC 1874
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 LysLeuGluAlaLysTyrValAlaAsnArgGlu...L 529
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1875 TTATTAAGCTTATTAATCTAAATGAAGAAAAATGCAATTAACACCGT 1924
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
529 euValLeuAsnAlaAlaGlyGlnGluAlaGlyAsnTyrThrVal 543
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1925 TGGTACACTGACAGCAGCGGCGAGCTGAGCGATTAAGTAAATGGGT 1974
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
544 ValLeuThrAlaLysSerGlyLysGluAla...LysAlaThr 557
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1975 TGGGTAGTATCAACCAAAACGTAAGCAAGAAAGAAAGCAATCAAGTTAA 2024
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
557 rLeuAlaLeuGluLeuLysAlaProGlyAlaPheSerLysPheGluVal 574
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2025 ACAAGCTGATGAAGTCTCTTACCGAGCC 2055
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
574 rGlyLysAspThrGlyLysAspLysTyrValThrGluGlnAsnGlnLys 590
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2056 GGTGCTGCTACGCTTACTTCCAAA...TCTGAAAACCGTTAA 2094
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
591 AsnAlaMetThrValSerValLeuProValAlaAsnGlyLeuValLe 607
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2095 CATACGATTACCGTTAGTGTGGCTGAACAAAGCGATTCGGCTCTGA 2144
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
607 uLysGlyAlaGluAlaAlaGluLeuLysValThrThrAsnLysGluG 624
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2145 AAAAGATGGCGATTAATTAAGCTCAAAAGTGAATATCAAAAC 2187
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
624 lLysGluValAlaAspAlaThrAspAlaGlnValThrValGlnAsnSer 640
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2188 ACTGATTAATGTTTAAACCTGTGCTGTAATAT 2217
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
641 ValIleThrValGlyGlnGlyAlaLysAlaGluTyrThrLysValThr 657
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2217
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
657 rValValLeuAspGlyLysLeuIleThrThrHisSerPheLysValVal 674
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2218 GGTACTGCTGTCACATAAAGGTGGCTTGAACCTGTA 2254
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
674 sPThrAlaProThrAlaLysGlyLeuAlaValGluPheThrSerThrSer 690
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

2255 AAACGTGAGCGACT.....GATGCAATCGCGGTAAAGTAACT 2292
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
691 LeuLysGluValAlaProAsn...AlaAspLeuLysAlaAlaLeuLys 706
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2293 GTAAAGATGCTACTGCTAATGACGCTGATAGAAAGTCGCACTGTAA 2342
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
706 nIleLeuSerValAspGlyValProAlaThrThrAlaLysAlaThrAla 723
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2343 AGATGTT.....GCAACCGCAATTAAATGTCGCGCGA 2374
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
723 eAsnValGluPheValSerAlaAspThrAsnValAlaAlaGlnAsnGly 739
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2375 CTTTGTGTAAGAACAGAGATTTACTACTCTATTGATGAAGATATCTT 2424
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
740 ThrValGlyAlaLys.....GlyAlaThrSerIleTyr 750
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2425 ACAGATTAACGGCAAAAGATGACGCACTTAACGGCGGCTTAACTTACCT 2474
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
750 rVal.....LysAsnLeuThrValValLysAspGlyLysGluGlnLysV 765
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2475 TAAAGCAGGTAAAGCTGAAGATTAAACGTGATGGAAGA....AATA 2518
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
765 aGluPheAspLysAlaValGlnValAlaValSerIleLysGluAlaLys 781
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2519 TTACTTTTGAAGCTTGGCGAAGAAC.....CTTGAGGTGAAGAACTGCGAAA 2562
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAA924BP9
seq_documentation_block:
ID AAA92499 standard; DNA: 7253 BP.
XX
AC AAA92499:
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) gene from type c strain API.
XX
KW Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
KW non-typable Haemophilus influenzae; anti-inflammation; auditory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen; ds.
XX
OS Haemophilus influenzae.
XX
PN WC200055191-A2.
XX
PD 21-SEP-2000.
XX
PE 16-MAR-2000; 2000WC-CA00289.
XX
PR 16-MAR-1999; 99US-0268347.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Loomore SM, Yang Y, Klein MH;
XX
WP1: 2000-618897/59.
XX
DR P-PSDB; AAB23860.
XX
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection.
XX
PS Claim 1; Fig 24; 275pp; English.
XX
CC The present sequence represents a Haemophilus influenzae adhesin
CC (Hia) gene from the type c Haemophilus influenzae strain API.
CC Hia genes and proteins have anti-inflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a

```

CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.

XX
 Sequence 7253 BP: 2495 A; 1338 C; 1782 G; 1638 T; 0 other;

alignment_scores:

Quality: 238.50 Length: 849
 Ratio: 0.575 Gaps: 42
 Percent Similarity: 48.881 Percent Identity: 22.968

alignment_block:

US-09-754-947-1 x AAA92499 ..

Align seg 1/1 to: AAA92499 from: 1 to: 7253

23 GluLysGlyAlaValLysGlyAsnAspLysGlyMetPheGluProGly 39
 |||||
 235 GAAAAAGAAAGTTACAGAAATTCAAATGGGAAATATTTCACACATTA 284
 39 sGluLeuThrArgAlaGluAlaAlaThrMetAlaGlnIleLeuAsnL 56
 | : : : : : ||| : : : : : ||| : : : : :
 285 AGGAGTACTAAAGACCGAGCAATCACCCTCAAGCGCGCAC..AAC 331
 56 euProIleAspLysAspAlaLys...ProSerPheAlaAspSerGlnGly 71
 || ||| : : : : : : : : : : : : : : : :
 332 TGAAATCAACAAAGCACCAATGCCAGTAGCTCACTACTCGCGAAA 381
 72 GlnTPryThrProPheIleAla...AlaValGluLysAlaGlyVal... 86
 : : : : : ||| : : : : : : : : : : : : : : :
 382 AAAGACCTCAGATCTGACCACTGTCACATGAAATTTATCGTTGG 431
 87
 432 CGCAAAACGCGATAAAGTTGATATTACAGTGTGCAATGGCTTGAAT 481
 96ProAsnGlyLysIle..... 100
 482 TGGCGAAACAGGTAAACGGAATGTTCAATTGATGTTGATTCACACT 531
 101AspArgValSerMetAlaSerLeuLeuValGluAlaTyrLysLe 115
 ||| : : : : : : : : : : : : : : : :
 532 TTGCCTGATGCGGTAAACGATACAGGTGTGTTAAGTTCTCAAGTTTAC 581
 115 uAspThrLysValAsnGlyThrProAlaThrLysPheLysAspLeuGluT 132
 : : : : : ||| : : : : : ||| : : : : :
 582 ACCTAATGATGTTGAAAAAACAAGAGCTGCACACTGTTAAAGATGTTTAA 631
 132 hLeuAsnTyrGlyLysGluLysAlaAsnIleLeuValGluLeuGlyIle 148
 : : : : : : : : : : : : : : : :
 632 ATGCGAGTTGGAAACATTTAAAGTGCTAAAACT..... 663
 149 SerValGlyThrGlyAspGlnTyrProGlyLysThrValThrLysAl 165
 : : : : : : : : : : : : : : : :
 664GCTGGAGGTATGTTGAGAGTGTGATTGATGTCGCGTTA 704
 165 aGluAlaAlaGlnPheIleAlaLysThrAspLysGlnPheGlyThrGluA 182
 : : : : : ||| : : : : : : : : : :
 705 TTAATATGTTGAATTTATTACAGCGCATAAACACGCTTGATGTTGAT 754
 182 lAlaLysValGluSerAlaLysAlaValThrThrGlnLysValGluVal 198
 : : : : : ||| : : : : : : : : : :
 755 TAAACGCTAAAGAAAAACGTTAAACA.....ACGCAAGTG 789
 199 LysPheSerLysAlaValGluLysLeuThrLysGluAspIleLysVal... 214
 ||||| : : : : : : : : : : ||| : : : : :
 1576 ACCAAAGATGGCATTTAAAGCAGGTAAATAAGCTAT.....AC 1613

790 AATTTCACACCGAAAAACCTCTGTATCAAAAGAAAAAGCGGTAACTTAT 839
 215 ThrAsnLysAlaAsnAsp.....LysValLeuValLysGluValT 229
 ||||| : : : : : ||| : : : : :
 840 TCTGGAAAGAGATTAACGACACAAATTAAGTTACAACTACAGCGGA 889
 229 hLeuSerGluAspLysArgSerAlaThrValGluLeuTyrSerAsnLeu 245
 || : : : : : ||| : : : : : ||| : : : : :
 890 CTGATATATCAAGATGAGGATATGGCTTAGTCACGCAAAAGCTGTGAT 939
 246 AlaAlaLysGlnThrTyrThrValAspValAsnLysValGly..... 259
 ||| : : : : : : : : : :
 940 GATGCT.....GTGAACAAGCTGTGTGAGAGT 968
 260 LysThrGluValAla.....ValGlyS 267
 ||||| : : : : : ||| : : : : :
 969 TAAACACACTACTGCTAATGTCAAATGGCACCTTCCCACTGTGGCT 1018
 267 erLeuGluAlaLysThrIleGluMetAlaAspGlnThrValAlaAsp 283
 || : : : : : ||| : : : : : ||| : : : : :
 1019 CAGCACAATGTAACTTTGAAAGTGGGATGTACAAACAGCTCA... 1065
 284 GluProThrAlaLeuGlnPheThrValLysAspGluAsnGlyThrGluVa 300
 1066GTAACTAAAGTACTTAACGGCAATGGCAT 1094
 300 lValSerPro.....GluGlyIleGluPheValThrP 311
 : : : : : : : : : : : : : : : :
 1095 CACTGTAAAGTACGACCGGAAAGTTGGGACGCGTTGAATTT.....G 1138
 311 roAlaAlaGluLysIleAsnAlaLysGlyGluIleThrLeuAlaLysGly 327
 : : : : : ||| : : : : : : : : : :
 1139 ATAGCGATAAAAAATC.....GTTGCAAGATACG 1167
 328 ThrSerThrThrValLysAlaValTyrLysLysAspGlyLysValValAl 344
 ||||| : : : : : ||| : : : : :
 1168 ACCGCCTTACTGTGACAGGT.....GTAAGGTAGCGTA 1202
 344 aGluSerLysGlu.....ValLysValSerAlaGluGluAlaVal 358
 : : : : : ||| : : : : : ||| : : : : :
 1203 AATGCTTAAGAGATGACAAAGAAAAAAGCTTGTAAAGCAGCGATTGG 1252
 358 AlaIleSerIleSerAsnTyrThrValAlaGluGlnAsnLysAlaAspHe 374
 ||||| : : : : : ||| : : : : : ||| : : : : :
 1253 TAACAGCTTTAGTAATCTAAGTTGAAAGCAAAAGCTGAGGCTCAT... 1299
 375 ThrSerLysAsp.....PheLysGlnAsnAsnLysValTyr 386
 ||||| : : : : : : : : : : : : : : : :
 1300 ACTGATCTGATGTCGCGCTGAGGGGATTTCAAAAGCAAGAGTCAA 1349
 386 rGluGlyAspAsnAlaTyrValGlnValGluLeuLysAspGlnPheAsnA 403
 ||||| : : : : : ||| : : : : : ||| : : : : :
 1350 AGCAGCGCAACGGA.....ACCTTAAAG 1375
 403 lAlaValThrThrGlyLysValGluTyrGluSerLeuAsnThrGluValAla 419
 : : : : : ||| : : : : : : : : : :
 1376 CGGCGAAGACTTAAAGTGAAGAACAGATGTGCGCACTTAACTATTCA 1425
 420 ValValAspLysAlaThrGlyLysValThrVal.....Le 431
 : : : : : ||| : : : : : : : : : :
 1426 CTGCAAGATGCTTTAAGCGGTTTAAAGAGCATTTAGTGGTGAAC 1475
 431 uSerAlaGlyLysAlaProValLysValThrValLysAsp..... 444
 : : : : : ||| : : : : : ||| : : : : :
 1476 TAAATGCGGAATATGATGCGAAACCGTATCAACAAACGTTTAACCA 1525
 445SerLysGlyLysAlaLeuValSerHisThrValGluIle 457
 : : : : : ||| : : : : : : : : : :
 1526 TCACGCGCAGAGGTATGCGGTACGACAGGTACAAACACATGACGCTA 1575
 458 GluAlaPheAlaGlnLysAla...MetLysAspIleLysLeuGluLysTh 473
 : : : : : ||| : : : : : ||| : : : : :
 1576 ACCAAAGATGGCATTTAAAGCAGGTAAATAAGCTAT.....AC 1613

[illegible]

```

FT      /function= transcription initiation site
FT      /note= "p3"
FT      misc_signal
FT      509..509
FT      /function= r
FT      /function= transcription initiation site
FT      /note= "p4"
FT      524..524
FT      /function= s
FT      /function= transcription initiation site
FT      /note= "p5"
XX
XX      JP03094683-A.
XX
XX      19-APR-1991.
XX
XX      06-SEP-1989; 89JP-0229304.
XX
XX      06-SEP-1989; 89JP-0229304.
XX
XX      (HIGE-) HIGETA SHOYU KK.
XX
XX      WPI: 1991-159801/22.
XX
XX      P-PSDB: AARI2083.
XX
XX      New protein HWP gene - comprising specified DNA sequences,
XX      permitting prodn. of large amts. of HWP for use as food protein
XX
XX      Disclosure: Fig 1; 8pp; Japanese.
XX
XX      Using this sequence HWP protein can be produced in large amounts by
XX      recombinant techniques. The protein is used in food. Promoters P1-
XX      P5 are very strong and can be used to express other proteins.
XX      According to the specification, triplet TTC (bases 548-550)
XX      encode amino acid MET.
XX
XX      Sequence 4330 BP; 1402 A; 806 C; 996 G; 1126 T; 0 other:

alignment_scores:
    Quality: 231.50      Length: 892
    Ratio: 0.550        Gaps: 45
    Percent Similarity: 47.197      Percent Identity: 21.076

alignment_block:
US-09-754-947-1 x AAQ11789 ..
Align seg 1/1 to: AAQ11789 from: 1 to: 4330

17 SerIleAsnTYrLeuValGluLysGlyAlaValLysGlyAsnAspLysG1 33
   ::::::::::: ||| ||| ||| ||| ::::
755 ACCGTTAAACGCTGTGAAGCTCTTGGCTGTAGCAGCGTTATGCGCAACGG 804
33 ymerPheGluProGlyLysGluLeuThrArgAlaGluAlaIleThrMetC 50
   | ||| ||| ||| ::::::::::: ||| ||| ||| ||| ||| ||| |||
805 CGATTTCGCTGCAGACAAACCATCCTCGTGCAGAGTTCGCTACTCTAA 854
50 eValaGluIleLeuAsnLeuProIleAspLysAspAlaLys..... 63
   ::::::::::: ::||| ::::::::::: ||| ||| ||| ||| ||| |||
855 TCGTTCGCGCTCGCGCTG.....GAGCAAGTGGCAAACTGCGCACAA 898
64 .....ProSerPheAlaAspSerGlnGly...GlnTPryrThrProPh 77
   ::::::::::: ||| ::::::::::: ||| ||| ||| ||| ||| |||
899 TTCAACACTACTTACACAGATGTAGATCCACTGCTGCTGCTGCTTT 948
77 eIleAlaIleValGluLysAlaGlyAlaIleLysGly...ThrGlyAsn 93
   |::: ::::::::::: ::::::::::: ||| ||| ||| ||| ||| |||
949 TGTAAACGTAGCTTCGCTGAAGAAATCGTAAAGGTTCCCGGCAAT 998
93 LypPheGluProAsnGlyLysIleAspArgValSerMetAlaSerLeu 109
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
999 CTTTAAACCAACAAACAGATTACTTATGCTGAAGCAAGTAAACATGATC 1048
110 ValGluAlaIleLysLeuAspThrLysValAsnGlyThrProAlaIleTh 126

```

```

||||:||||
1049 GTTCCTGCTTGGTTATGAGCCATCCGTTGAAGTGTA..... 1087
126 sPheLysAspLeuGluThrLeuAsnTrpGlyLysGluLysAlaAsnIleL 143
1088 .....TGGCCGACACGATGATCTCAAG 1112
143 euValGluLeuGlyIleSerValGlyThrGlyAspGlnTrpLupProLys 159
   |||||:||||: |||
1113 GTTCGCACTGACATGCAAAAGT..... 1138
160 LysThrValThrLysAlaGluAlaAlaGlnPheIleAlaLysThrAspLy 176
   ::::::::::: ::| ||||| ||| |||
1139 .....ATCACACACCCTAACATGCAGCAGCTTCGCGGCAATCTTCAA 1182
176 sGlnPheGlyThrGluAlaAlaLysValGluSerAlaLysAlaValThrT 193
   | |||:||||: |||
1183 A.....ATGCTGG 1190
193 hrGluLysValGluValLysPheSerLysAlaValGluLysLeuThrLys 209
   ::|: |||||: ||| ||| ||| |||
1191 ACAACGCTCTTCGCGTTAAGCTGATGAGCAAAATCGAATACGCTACTGAC 1240
210 GluAspIleLysValThrAsnLys.....AlaAsnAsnAspLysVa 223
   ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1241 ATCCGTTTAAACGTAACTGACGAAACCTCTTGACTTAATATTGAAAGT 1290
223 lLeuValLysGluValThrLeuSerGluAspLys.....ArgSerAlaT 238
   | |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
1291 TACCTTACGTGATATGACTGGGCTCACGAAAGGCTAACATTTCTCATG 1340
238 hrValGluLeuThrSerAsnLeuAlaAlaLysGlnThrTyThrValAsp 254
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1341 AATGCACTGTGTACAAACGTAACCTGCTATGCTGGGTGATTGAAA 1390
255 ValAsnLysValGlyLysThrGluValAlaValGlySerLeuGluAlaLy 271
   ::|||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
1391 GCAAAATGAAGTT.....ACTTGAATGAANA 1416
271 sThrIleGluMetAlaAspGlnThrVal.....ValAlaAsp.....G 284
   | |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1417 AGATCTGATCTGGGTACCAACACTACTTATTAAGTACTGTAAGCATCA 1466
284 lupProThrAla.....LeuGlnPheThrValLysAspLup 295
   ::|||:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1467 ATCCTAACGCAATTGATGTCAAAAGTACAAAGTGAATCAAAAGATGAC 1516
296 AsnGlyThrGluValAlaSerProGluGlyIleGluPheValThrProAl 312
   ::|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1517 CGAGAAATGTCAATCGTTTGATGGAAGTTCGGAAGCAGAAATGTGCT 1566
312 aAlaGluLysIleAsnAla.....LysGlyValThrLeuAlaL 326
   ::|: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1567 TATGACCGCTGAGTCTCTGTACTTCAAAAGTAAACCTTC..... 1609
326 ysglyThrSerThrThrValLysAlaValTyLysLysAspGlyLysVal 342
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1610 .....ACAGATGATATTGTAAAGATCTTACAAAGTGTGAT..... 1645
343 ValAlaGluSerLysGluValLysValSerAlaGluGlyAlaAlaValAl 359
   ::|||:||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1646 .....TTGATGATGTAATAATGCAAAATGTGATG..... 1675
359 aSerIleSerAsnTrpThrValAlaGluLysAsnLys..... 371
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1676 .AGCGAAGAAATCGTATCGCTTACGAGCAAGATACGAAATCACTTAACT 1724
372 .....AlaAspPheThrSerLysAspPheLysGln 381
   ::|||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1725 TCACTCGTTTCAACATCCAGTACGATGCTCTGAGCAAAATTTACAAAGC 1774
382 AsnAsn.....LysValTyGluGlyAspAsn.....Al 391
   |||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


92x2